Time Series Modelling Version 4.47

Models and Methods

James Davidson

University of Exeter

22nd September 2015

Contents

1 Introduction 6
1.1 Copyright Notice 6
1.2 Disclaimer 6
1.3 Acknowledgements 6
2 Linear Regression 7
2.1 Regressor 'Types' 7
2.2 Instrumental Variables 7
2.3 Restrictions and Tests 7
2.4 Cointegrating Regressions 8
2.5 System Estimation 8
3 Panel Data 9
3.1 Data Transformations 9
3.2 Dummy Variables 9
3.3 Estimation Methods 10
3.3.1 Ordinary Least Squares for Fixed Effects 10
3.3.2 Generalized Least Squares for Random Effects 10
3.3.3 Maximum Likelihood for Random Effects 11
3.3.4 Instrumental Variables/GMM 11
3.4 Tests and Diagnostics 11
3.5 System Estimation 12
4 Single Equation Dynamic Models 13
4.1 Linear Models of the Conditional Mean 13
4.2 Definitions and Details 14
4.2.1 Lag Polynomials 14
4.2.2 Fractional Difference Operator 14
4.2.3 Intercept and Linear Trend Dummies 14
4.2.4 Regressor Types 15
4.2.5 Inequality Constraints 15
4.2.6 Polynomial Distributed Lags 15

©James Davidson 2015

4.3 Bilinear Models 16
4.4 Nonlinear Moving Average Models 16
4.5 Error Correction Models 17
4.5.1 Equilibrium Relations 17
4.5.2 Fractional Cointegration 18
4.5.3 Nonlinear Error Correction and Nonlinear AR 18
4.6 User Coded Functions 18
4.6.1 Coded Formulae 18
4.6.2 Formula Types 19
4.6.3 Coded Equations 19
4.6.4 Coded Component 20
4.6.5 Coded Error Correction Mechanism. 20
4.6.6 Coded Moving Average Model 20
4.6.7 Coded Equilibrium Relations 20
4.6.8 Ox Coding 20
4.7 Discrete Data Models 21
4.7.1 Probit and Logit Models 21
4.7.2 Ordered Probit and Logit 21
4.7.3 Count Data 22
4.7.4 Autoregressive discrete models 22
4.7.5 Zero-inflated Poisson and ordered Probit 22
5 Systems of Equations 23
5.1 The Basic Model 23
5.2 Definitions and Details 23
5.2.1 System Notation 23
5.2.2 System Exogenous Variables 23
5.2.3 Simultaneous Equations 23
5.2.4 Nonlinear Systems 23
5.3 Error Correction Models 24
6 Conditional Heteroscedasticity Models 26
6.1 Single Equation Models 26
6.2 Definitions and Details 26
6.2.1 GARCH Parameterization 26
6.2.2 HYGARCH and FIGARCH 27
6.2.3 Asymmetric GARCH and Power GARCH 27
6.2.4 GARCH Regressors 27
6.2.5 EGARCH 27
6.2.6 HYEGARCH and FIEGARCH 28
6.3 Conditional Heteroscedasticity in Discrete Data Models 28
7 Conditionally Heteroscedastic Systems 29
7.1 Implemented Model Variants 29
7.2 Definitions and Details 30
7.2.1 Multivariate GARCH 30
7.2.2 DCC Multivariate GARCH 30
7.2.3 BEKK Multivariate GARCH 30
8 Regime Switching 31
8.1 Simple Markov switching 31
8.2 Explained Switching 32

©James Davidson 2015

8.3 Hamilton's Markov switching model 32
8.4 The Smooth Transition (ST) Model 33
8.4.1 Single Transition Model 33
8.4.2 Double Transition Model 34
8.4.3 Structural Change (Date-Transition) Model 34
8.5 Testing for Breaks and Regimes 34
9 Post-Estimation Options 35
9.1 Residuals and Associated Series 35
9.2 Model Performance and Selection Criteria 36
9.3 Q Tests 37
9.4 Ex-post forecasts 37
9.5 Ex-ante Multi-step Forecasts 38
9.5.1 Analytic Forecasts 38
9.5.2 Moving Average Coefficients (Impulse and Step Responses) 39
9.5.3 Forecast Error Variance Decomposition 39
9.5.4 Forecasting Regime-Switching Models 39
9.5.5 Monte Carlo Forecasts 40
10 Estimation Criteria 41
10.1 Single Equation Methods 41
10.1.1 Least Squares 41
10.1.2 Instrumental Variables 41
10.1.3 Gaussian maximum likelihood 41
10.1.4 Student-*t* maximum likelihood 41
10.1.5 Skew-student maximum likelihood 41
10.1.6 GED maximum likelihood 42
10.1.7 Whittle maximum likelihood 42
10.1.8 Probit and Logit 42
10.1.9 Ordered Probit and Logit 43
10.1.10 Poisson 43
10.1.11 Zero-inflated Discrete 43
10.2 System Methods 43
10.2.1 Least Generalized Variance 43
10.2.2 Generalized Method of Moments 44
10.2.3 Gaussian ML with Conditional Heteroscedasticity 44
10.2.4 Student-*t* ML with Conditional Heteroscedasticity 44
10.2.5 GED ML with Conditional Heteroscedasticity 45
10.3 Markov Switching Models 45
11 Standard Errors and Covariance Matrix Formulae 46
11.1 Information Matrix Formulae 46
11.2 Robust Formulae 46
11.3 HAC Variance Estimators 46
11.3.1 Bandwidth Selection 46
11.3.2 Pre-whitening 47
11.4 KVB Inconsistent Variance Estimates 48
12 Test Statistics 49
12.1 Wald Tests 49
12.2 Lagrange Multiplier Tests 49

©James Davidson 2015

12.3 Moment and Conditional Moment Tests 49
12.4 Information Matrix Test 51
12.5 Nyblom-Hansen Stability Tests 51
12.6 Andrews Structural Change LM Test 52
12.7 Consistent Specification Tests 52
12.7.1 Bierens Tests 52
12.7.2 Score-contribution Tests 53
12.7.3 Dynamic Specification Tests 53
12.8 Vuong's Test of Non-Nested Models 54
12.9 Cusum of Squares Test 54
12.10 'Sup-F' Tests 55
12.11 Tests of Integration Order 55
12.11.1 Tests of I(0) 55
12.11.2 Tests of I(1) 56
12.12 Bootstrap Test of I(0) 56
13 Simulation and Resampling Options 58
13.1 Monte Carlo Experiments 58
13.2 Parallel Processing 58
13.3 Numerical Test Distributions 58
13.4 Bootstrap Inference 59
13.5 Resampling Methods 61
13.5.1 Likelihood Model 61
13.5.2 Gaussian 61
13.5.3 Stable 61
13.5.4 Formula 61
13.5.5 Simple/Block Bootstrap 61
13.5.6 Stationary Block Bootstrap 61
13.5.7 Wild Bootstrap 62
13.5.8 Fourier Bootstrap 62
13.5.9 Sieve-AR Bootstrap 62
13.5.10 Data Resampling 62
13.6 Data Generation 63
13.6.1 Dynamic Data Simulation 63
13.6.2 The Static Bootstrap 64
13.7 Nonlinear Models 64
13.8 Panel Data 64
13.9 Subsampling Inference 65
13.10 The Fast Double Bootstrap 65
13.11 Warp-speed Monte Carlo for Bootstrap Estimators 66
14 Additional Features 67
14.1 Semiparametric Long Memory 67
14.1.1 Geweke Porter-Hudak Log-Periodogram Regression 67
14.1.2 Moulines-Soulier Log-Periodogram Regression 67
14.1.3 Local Whittle ML 67
14.2 Cointegration Analysis 68
14.3 Automatic Model Selection 69
14.3.1 ARMA Order Selection 69

 ©James Davidson 2015

14.3.2 Regressor Selection....69
14.4 SsfPack State Space Modelling 70
14.5 Calculator and Matrix Calculator 70
15 References 72
Index 78

©James Davidson 2015

1 Introduction

This document describes the econometric models that can be estimated in Time Series Modelling 4.47, and gives details of the forecasts, tests and other post-estimation options, optimization criteria, simulation and bootstrap options, and supplementary capabilities. It does not explain how to use the program. The Users' Manual is included in PDF format (`tsm4ghp.pdf`) as well as being available interactively via the program's Help system. The Appendices to this document (`tsm4app.pdf`) explain how to install and customize the program. The Programming Manual (`tsm4prg.pdf`) explains how to call the program functions from within a user's Ox program.

1.1 Copyright Notice

Time Series Modelling 4.47 is copyright © James Davidson 2002-2015, http://www.timeseriesmodelling.com/.
Please cite Time Series Modelling 4.46 in any publications where results obtained with the program are reported.
Ox 7.01 or later versions (© J. A. Doornik, 1994–2013) is required to run the package.
Ox Console is free to academic users only, from http://www.doornik.com. It should also be cited in any publications, please visit the web site for details.
Ox Professional is required for 64-bit installation.
The GUI version incorporates the following freely distributed copyright components:
OxJapi Version 2 ©2008-2013, Timothy Miller.
OxJapi ©2002, Christine Choirat, Rafaello Seri and Licensed under Gnu Lesser General Public Licence Version 2.1, February 1999 http://www.tinbergen.nl/~cbos/index.html
GnuDraw 6.3 Charles Bos http://www.tinbergen.nl/~cbos/index.html
GnuPlot 4.2.6 ©1986-1993, 1998-2008, Thomas Williams, Colin Kelley and others
http://www.gnuplot.info/

1.2 Disclaimer

This program is distributed with no warranties as to fitness for any purpose. Use it at your own risk.

1.3 Acknowledgements

Special thanks to Tim Miller for developing the OxJapi 2 package to run the GUI under the latest Java implementation; to Charles Bos for his estimable support with implementing and developing Gnudraw, and with the Linux implementation; to Andreea Halunga for her important programming contributions, especially in connection with analytic derivatives; to Paulo Dias Costa Parente for advice on teaching applications; Andrea Monticini for his many contributions to the development of TSM; to Ossama Mikhail for his initiative in setting up a discussion list; and all those TSM users, too numerous to mention, who have contributed helpful suggestions and bug reports.

2 Linear Regression

The Linear Regression dialog, accessed with the button on the toolbar, offers ordinary least squares (OLS), two-stage least squares (2SLS) and, for systems of equations, three stage least squares (3SLS) and seemingly-unrelated regressions (SUR). For example, VARs can be easily specified in the latter mode, with the specified set of lags generated automatically for each variable. All these estimators are computed in one or two steps, from closed formulae.

2.1 Regressor 'Types'

The components of a linear model can include intercept, trend, and regressors which can be allocated one of two Types. The main purpose of Types is to allow different orders of lags to be specified. By setting the scroll-bar in the Linear Regression dialog, lags can be included automatically without having to create them ahead of time, and different orders can be specified for each Type.

For example, dummy variables that should not be lagged can be entered as Type 1 variables. Type 2 variables are special, since this set can include the dependent variable (or variables, in a multiple equation model). In this case, the current value of the variable is automatically suppressed, and lags start from 1. For variables not in the 'dependent' set, lags start from zero. This makes it easy to include the lagged dependent variable as a regressor. While the maximum lag is set with the scrollbar, note that the specification can be "fine-tuned" by restricting individual coefficients to 0. This is done by checking the 'Fixed' checkboxes in the Values / Equation dialog.

2.2 Instrumental Variables

When 2SLS is selected an instrument set must be specified, which should include any exogenous variables in the equation. Lags of the additional instruments will be included to match those specified for the Type 2 regressors. Lagged endogenous variables are not used as instruments by default, but this is a selectable option.

2.3 Restrictions and Tests

It is possible to constrain coefficients to any chosen values, either zero or nonzero. This feature allows the imposition of identifying restrictions in simultaneous systems, for example. Tests of coefficient restrictions, based on the unrestricted regression, can be computed using the Wald principle. Zero restrictions, linear restrictions and nonlinear restrictions can be tested. In the latter case the restrictions are coded as algebraic expressions, allowing any degree of flexibility.

Diagnostic tests based on the Lagrange multiplier and conditional moment principles can be computed using either preset specifications (autocorrelation, RESET, heteroscedasticity and neglected ARCH) as well as user-selected test variables. Other options include the ADF and Phillips-Perron tests for cointegration, the Durbin-Watson statistic, and a general test of model significance, allowing for dummies and lagged dependent variables under the null hypothesis. By dividing the sample into estimation and forecast periods, Chow's forecasting and parameter stability tests can be computed. Advanced users also have the option of bootstrap methods, including bootstrap confidence intervals, test *p*-values, and bias corrections.

©James Davidson 2015

2.4 Cointegrating Regressions

To estimate cointegrating relations semi-parametrically, two methods are implemented; the Phillips-Hansen (1990) fully modified least squares estimator; and the least squares estimator augmented by lags and leads of the differenced right-hand side variables, as proposed by Saikkonen (1991) and Stock and Watson (1993). A range of bandwidth, kernel and automatic lag/lead selection options are provided to support these procedures.

The augmented Dickey Fuller and Phillips-Perron tests for the null hypothesis of cointegration are also implemented in this dialog. By regressing a variable on intercept or intercept and trend only, these options can also be used to test the I(1) hypothesis. See also Section 12.8.2 for alternative options for I(1) tests.

2.5 System Estimation

A system of equations is specified by a selecting a set of dependent variables, and a set of explanatory variables. The latter are allocated to one of the two Types, each with its own lag length. The same explanatory variables appear in each equation, by default, but variables can be optionally excluded from an equation by "fixing" their coefficients to 0. The Type 2 set can include dependent variables, in which case the zero-order lags are automatically suppressed.

To estimate the system by 3SLS requires only that a set of instruments be selected. Additional instruments are included as lags up to the maximum specified by the Type 2 regressors setting. The Type 1 regressor set can include both current and lagged dependent variables. The current values are automatically excluded from the equation in which they appear on the left-hand side. Of course, identifying restrictions may also need to be imposed by hand, equation by equation.

©James Davidson 2015

3 Panel Data

Panel regression is supported when the data file is created with a specified format.

The basic model is assumed to take the form

$$y_{it} = \beta' \boldsymbol{x}_{it} + \lambda_t + \eta_i + v_{it}, \quad t = T_{1i}+1,\ldots,T_{2i},\ i = 1,\ldots,N \tag{3.1}$$

The subscript *i* indexes *individuals*, or cross-sectional units, while *t* indexes *dates*. v_{it} is a disturbance with mean 0 and variance σ_v^2, and distributed independently of $\boldsymbol{x}_{it}$ for all *i* and *t*. Panels can be unbalanced, with different start and end dates for different individuals, and can even be irregular with missing time periods, although this possibility is not indicated explicitly in the notation of equation (3.1). Except in the irregular case, dates can be seasonal, with a 'year + quarter' or 'year + month' format.

The vector $\boldsymbol{x}_{it}$ ($p \times 1$) contains regressors which can consist of current and lagged exogenous variables, and also lagged endogenous variables. It can optionally include an intercept and also a trend term (i.e., $x_{1,it} = t - \min_i T_{1i}$). We define $T_i = T_{2i} - T_{1i} + 1$ for brevity, and let the total number of observations be $O = \sum_{i=1}^{N} T_i$.

3.1 Data Transformations

The variables $\boldsymbol{w}_{ti} = (y_{it}, \boldsymbol{x}_{it}')'$ may be automatically subjected to one of the following transformations, denoted generically by $\boldsymbol{w}_{it}^*$.

1. *Deviations from time means* : $\boldsymbol{w}_{it}^* = \boldsymbol{w}_{it} - \bar{\boldsymbol{w}}_i$ where $\bar{\boldsymbol{w}}_i = T_i^{-1} \sum_{t=T_{1i}+1}^{T_{2i}} \boldsymbol{w}_{it}$.

2. *Time means*: $\boldsymbol{w}_{it}^* = \bar{\boldsymbol{w}}_i, t = T_{1i}+1,\ldots,T_{2i}$ where $\bar{\boldsymbol{w}}_i = T_i^{-1} \sum_{t=T_{1i}+1}^{T_{2i}} \boldsymbol{w}_{it}$. Thus, there are *O* "observations", but because of the repeated entries the true degrees of freedom for a regression with these data is $N - p$. If $\lambda_t = 0$, the averaged form of equation (3.1) has a disturbance of the form $\eta_i + \bar{v}_i$.

3. *Time differences*: $\boldsymbol{w}_{it}^* = \Delta \boldsymbol{w}_{it} = \boldsymbol{w}_{it} - \boldsymbol{w}_{i,t-1}$ for $t = T_{1i}+2,\ldots,T_{2i}$ and $\boldsymbol{w}_{T_{1i}+1,i}^* = \boldsymbol{0}$.

4. *Orthogonal deviations*: $\boldsymbol{w}_{it}^* = \left(w_{it} - \dfrac{\sum_{s=t+1}^{T_{2i}} w_{is}}{T_{2i} - t} \right) \left(\dfrac{T_{2i} - t}{T_{2i} - t + 1} \right)^{1/2}$ for $t = T_{1i}+1,\ldots T_{2i}-1$ and $\boldsymbol{w}_{iT_{2i}}^* = \boldsymbol{w}_{iT_{2i}}$. This transformation corresponds to the operations of differencing and then applying the appropriate transformation for GLS estimation of the equation, assuming that Δv_{it} is the difference of an independent sequence. (see Arellano 2003, page 17).

Under transformations 1, 3 or 4 applied to the variables in equation (3.1), note that η_i disappears.

3.2 Dummy Variables

Dummy variables can be generated and added to the equation automatically. The available options are as follows.

©James Davidson 2015

1. Individual dummies $d_{ti}^{I}(j) = 1, i = j, 0$ otherwise . Including these dummies in the equation for $j = 1,\ldots,N$ allows the estimation of the coefficients η_j (assuming the data have not already been transformed to remove these effects). This option is only available if no data transformation is selected.
2. Time dummies, $d_{ti}^{T}(s) = 1, t = s, 0$ otherwise , for $s = \min_i T_{1i} + 1, \ldots \max_i T_{2i}$. Include these dummies to estimate fixed time effects λ_s .
3. Group dummies, $d_{ti}^{G}(k) = 1, i \in G_k, 0$ otherwise where $\{G_k, k = 1,\ldots, M\}$ represents a partition of the individual indices into subsets.
   Group membership has to be specified by setting up indicator variables in the data set. This option can be used to estimate the model subject to the restriction that $\eta_i = \eta_j$ when $i, j \in G_k$.

If an intercept is included in $\boldsymbol{x}_{it}$, the dummies $d_{ti}^{I}(1)$ and $d_{ti}^{T}(1)$ are automatically excluded, to avoid the dummy variable trap.

3.3 Estimation Methods

Three estimation methods are implemented: ordinary least squares (OLS), two-step feasible generalized least squares (FGLS), and maximum likelihood.

3.3.1 Ordinary Least Squares for Fixed Effects

In OLS, the η_i terms are treated as fixed, and estimated as the coefficients of the dummies of type 1 ("fixed effects"). Note that the regression with data transformation of type 1 is identical to the regression with untransformed data including dummies of type 1, but in the second case the estimates of η_i are reported.

3.3.2 Generalized Least Squares for Random Effects

In the random effects model the η_i terms are treated as random variables with mean 0 and variance σ_η^2, and assumed to be distributed independently of $\boldsymbol{x}_{it}$ for all *i* and *t*. Therefore, the disturbances $\varepsilon_{it} = \eta_i + v_{it}$ are correlated, and efficient estimation requires a GLS regression. Letting $\tau = \sigma_\eta^2 / \sigma_v^2$, and assuming this known, the exact GLS estimator is obtained by computing the regression in the transformed variables $\boldsymbol{w}_{it}^{\tau} = \boldsymbol{w}_{it} - \theta_i \overline{\boldsymbol{w}}_i$ where

$$\theta_i = 1 - \left(\frac{\sigma_v^2}{\sigma_v^2 + T_i \sigma_\eta^2} \right)^{1/2} = 1 - \left(1 + T_i \tau\right)^{-1/2}. \tag{3.2}$$

In FGLS, τ is estimated from the residuals of the "within" and "between" regressions. These are, respectively, regressions in the individual mean deviations and the individual time means. In the latter regression the disturbances are $\overline{u}_i = \overline{v}_i + \eta_i$ with variances depending on *i*, $\sigma_i^2 = T_i^{-1}\sigma_v^2 + \sigma_\eta^2$. Letting

$$\hat{\sigma}_v^2 = (O - N)^{-1} \sum_{i=1}^{N} \sum_{t=T_{1i}+1}^{T_{2i}} (y_{it}^* - \hat{\boldsymbol{\beta}}^{*\prime} \boldsymbol{x}_{it}^*)^2 \tag{3.3}$$

and

©James Davidson 2015

$$\hat{\sigma}_0^2 = N^{-1}\sum_{i=1}^{N} T_i(\bar{y}_i - \hat{\bar{\boldsymbol{\beta}}}'\bar{\boldsymbol{x}}_i)^2 \tag{3.4}$$

denote the residual variances from the regressions under transformations 1 and 3 respectively, σ_η^2 is estimated by

$$\hat{\sigma}_\eta^2 = (\hat{\sigma}_0^2 - \hat{\sigma}_v^2)\frac{N}{O}. \tag{3.5}$$

Note that in the case of a balanced panel with $T_i = T$, $N / O = 1/T$. The second step of FGLS is performed by replacing τ by $\hat{\tau} = \hat{\sigma}_\eta^2 / \hat{\sigma}_v^2$ in (3.2).

Note that time effects cannot be modelled as random in this release.

3.3.3 Maximum Likelihood for Random Effects

Maximum likelihood extends the GLS approach by optimizing the concentrated Gaussian log-likelihood of the sample with respect to τ. This requires just a univariate numerical maximization (line search) over τ values. The criterion function is

$$L_T = -\frac{O}{2}\Big(\log(2\pi) + \log(\hat{\sigma}_v^2(\tau))\Big) - \frac{1}{2}\sum_{i=1}^{N}\log(1 + T_i\tau) \tag{3.6}$$

where $\hat{\sigma}_v^2(\tau)$ is defined similarly to (3.3), except that $(y_{it}^*, \boldsymbol{x}_{it}^*)$ are replaced by $(y_{it}^\tau, \boldsymbol{x}_{it}^\tau)$ as in (3.2).

3.3.4 Instrumental Variables/GMM

Panel data models may be estimated by IV if valid instruments are constructed by the user. Note that specialized GMM procedures for dynamic panel data models, unless they can be implemented by a suitable construction of the instrument series, are not implemented in this release.

3.4 Tests and Diagnostics

Two options are available for computing the covariance matrix for *t* values and Wald statistics.

1. Standard formula:

$$\boldsymbol{V}_1 = s^2\left(\sum_{i=1}^{N}\sum_{t=T_{1i}+1}^{T_{2i}} \boldsymbol{x}_{it}^*\boldsymbol{x}_{it}^{*\prime}\right)^{-1} \tag{3.7}$$

where $s^2 = (O-p)^{-1}\sum_{i=1}^{N}\sum_{t=T_{1i}+1}^{T_{2i}} \hat{v}_{it}^2$.

2. Robust formula:

$$\boldsymbol{V}_2 = \left(\sum_{i=1}^{N}\sum_{t=T_{1i}+1}^{T_{2i}} \boldsymbol{x}_{it}^*\boldsymbol{x}_{it}^{*\prime}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=T_{1i}+1}^{T_{2i}} \hat{v}_{it}^2\boldsymbol{x}_{it}^*\boldsymbol{x}_{it}^{*\prime}\right)\left(\sum_{i=1}^{N}\sum_{t=T_{1i}+1}^{T_{2i}} \boldsymbol{x}_{it}^*\boldsymbol{x}_{it}^{*\prime}\right)^{-1} \tag{3.8}$$

The following test statistics are reported automatically, under the setup indicated.

1. Jarque-Bera test for normality of "within" disturbances. (All cases.)

©James Davidson 2015

2. Breusch and Pagan (1980) LM test of the null hypothesis $\sigma_\eta^2 = 0$. (OLS without transformations.)
3. Bhargava, Franzini and Narendranathan (1982) modified Durbin-Watson statistic. This tests the null hypothesis of serial uncorrelatedness of "within" disturbances. (Not under transformation 3.)
4. Hausman (1978) test for correct specification in the random effects model. Under the null hypothesis, $E(\eta_i \mid \boldsymbol{x}_{iT_1},\ldots,\boldsymbol{x}_{iT_2}) = 0$. (FGLS and ML).

3.5 System Estimation

Systems of panel equations with fixed effects can be estimated in the same way as for one-dimensional samples. Either least squares (SUR) or instrumental variables can be specified. Systems with random effects cannot be estimated in this release.

©James Davidson 2015

4 Single Equation Dynamic Models

The Dynamic Equation dialog, accessed with the button on the toolbar, gives access to models that require optimizing a log-likelihood function or other criterion function, although linear models can also be estimated. The options include ARMA, ARIMA and ARFIMA models, and error-correction (ECM) models. Conditional variance models (see Section 6) include ARCH, GARCH, and numerous variants. For systems of equations, the vector generalizations of all these models are available (see Sections 5 and 7). FIML for linear simultaneous systems is implemented automatically if current endogenous variables are included as explanatory variables. All the specifications allow stochastic regime-switching in mean and variance, including Markov switching and smooth transition (see Section 8).

4.1 Linear Models of the Conditional Mean

Let Y_t for $t = 1,\ldots,T$ denote the time series to be modelled. Consider the ARFIMA class of dynamic regression models, having the general form

$$\phi(L)(1-L)^{d_1}(Y_t - \gamma_{01} - \gamma_1 t - \boldsymbol{\pi}_1'\boldsymbol{x}_{1t}) = \gamma_{02} + \boldsymbol{\pi}_2'\boldsymbol{x}_{2t} + \theta(L)v_t \tag{4.1}$$

where

$$v_t = \boldsymbol{\pi}_3'\boldsymbol{x}_{3t} + u_t \tag{4.2}$$

and at most one of γ_{01} and γ_{02} can be different from zero. The $\boldsymbol{x}_{jt}$ for $j = 1,2$ and 3 are vectors of explanatory variables, entering with coefficient vectors $\boldsymbol{\pi}_j$.

Equation (4.1) encompasses a range of options of which no more than a few are likely to be selected at once. The most basic time series model for a single series is the univariate ARMA(p,q) form. Setting $d_1 = 0$ and suppressing the explanatory variables, this might be written as

$$\phi(L)Y_t = \gamma_{02} + \theta(L)u_t \tag{4.3}$$

in which $\phi(L)$ and $\theta(L)$ represent the autoregressive (AR) and moving average (MA) lag polynomials, of order p and q respectively. For example, if $p = q = 1$ the equation would have the explicit form

$$Y_t = \gamma_{02} + \phi_1 Y_{t-1} + u_t - \theta_1 u_{t-1}. \tag{4.4}$$

Note that the specification

$$Y_t - \gamma_{01} = \phi_1(Y_{t-1} - \gamma_{01}) + u_t - \theta_1 u_{t-1} \tag{4.5}.$$

is equivalent, implying the identity $\gamma_{02} = \phi(1)\gamma_{01}$. Either form can be estimated.

By setting $d_1 = 1$ a unit root can be imposed, defining the nonstationary ARIMA(p,1,q) model. This is equivalent to differencing the series before fitting the ARMA model, and so takes the form

$$\phi(L)\Delta Y_t = \theta(L)u_t. \tag{4.6}$$

In the ARFIMA model, the simple difference is replaced by a fractional difference. This is equivalent to expressing the left-hand side of the equation as

$$\phi(L)(1-L)^{d_1}Y_t = \theta(L)u_t. \tag{4.7}$$

©James Davidson 2015

This model reduces to the ARMA model if $d_1 = 0$ and the ARIMA model if $d_1 = 1$. It is stationary and invertible if $|d_1| < 0.5$. See Section 4.2.3 for details of how to include mean and trend components to these models.

4.2 Definitions and Details

4.2.1 Lag Polynomials

$\phi(L)$, $\theta(L)$, are finite-order lag polynomials, such that, for example,

$$\phi(L) = 1_t - \phi_1 L - \cdots - \phi_p L^p \tag{4.8}$$

where L denotes the lag operator, such that $Lx_t = x_{t-1}$. Thus,

$$\phi(L)Y_t = Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p}.$$

By default, the values ϕ_j and θ_j are reported.. However, the sign convention can be optionally changed for the MA coefficients, reporting $-\theta_j$ instead.

4.2.2 Fractional Difference Operator

This is

$$(1-L)^d = \sum_{j=0}^{\infty} b_j L^j \tag{4.9}$$

where $b_0 = 1$ and

$$b_j = \frac{-d\Gamma(j-d)}{\Gamma(1-d)\Gamma(j+1)} = \frac{j-d-1}{j} b_{j-1}, \quad j \geq 1. \tag{4.10}$$

The infinite lag structures represented by (4.9) are approximated in the sample by replacing $(1-L)^d$ by

$$(1-L)_+^d = \sum_{j=0}^{t-1} b_j L^j . \tag{4.11}$$

In other words, the lag distribution is truncated at the beginning of the available sample.

In long memory models the omission of the pre-sample observations can change the distribution of the estimates, even asymptotically. A technique for correcting this effect is implemented experimentally; see Davidson and Hashimzade (2009).

4.2.3 Intercept and Linear Trend Dummies

These are built-in options, and do not need to be added as dummy regressors. γ_{01} or γ_{02} in equation (4.1) are called, respectively, intercepts of Type 1 and Type 2. At most one can be present. In autoregressive models the 'Type' chosen makes no difference to the fit, but the value and interpretation of the coefficient is different, compare equations (4.4) and (4.5). The built-in trend is of Type 1, with coefficient γ_1. A Type 2 trend dummy is not built in but it could be included as a generated regressor in the set $\boldsymbol{x}_{2t}$ in equation (4.1). If included in equation (4.4), for example, its coefficient would have the form $(1-\phi_1)\gamma_1$. Also note that a Type 2 intercept would become $\gamma_{02} + \phi_1\gamma_1$ in this case.

If a unit root is imposed, as in equation (4.6), γ_{01} is unidentified. The Type 1 trend coefficient γ_1 becomes in effect an intercept, behaving exactly like a Type 2 intercept. You cannot add both of these simultaneously!

©James Davidson 2015

In the fractional model too, the different types of intercept represent different models, not just alternative parameterizations. While $\gamma_{01} \neq 0$ represents merely a location shift of the fractionally integrated process, $\gamma_{02} \neq 0$ implies the presence of a deterministic trend of $O(t^d)$. When $d < 0$, this term degenerates to 0 after a finite number of steps. γ_{02} would be asymptotically unidentified in that case, and should be suppressed.

In the bilinear model (see Section 4.3) the two intercept cases likewise represent different models, with different dynamics.

4.2.4 Regressor Types

The variables $\boldsymbol{x}_{1t}$, $\boldsymbol{x}_{2t}$ and $\boldsymbol{x}_{3t}$ in equation (4.1) are referred to as regressors of Type 1, Type 2 and Type 3, respectively. As in the linear regression, each Type can be specified with a different number of lags. However, in an ARMA or ARFIMA equation, the Types have special significance because of the way the variables enter the dynamics in equation (4.1):

- A model with only Type 1 regressors can be thought of as exhibiting "error dynamics", since a transformation allows it to be rewritten with only the error term u_t entering in lagged form.
- A model with Type 2 regressors exhibits "structural dynamics", since it has a distributed lag representation. As in a linear regression, the dependent variable can enter as a Type 2 variable with lags, and in this case, the current value is suppressed. However, specifying an AR form is a more natural way to include the lagged dependent variable in most situations.
- Type 3 regressors act in effect as components of the error term, adjusting its mean systematically.

Important note: when a unit root is imposed, the effect on equation (4.1) is to replace Y_t by ΔY_t, t by 1 and $\boldsymbol{x}_{1t}$ by $\Delta\boldsymbol{x}_{1t}$. However, the Type 2 and 3 regressors enter as before.

4.2.5 Inequality Constraints

Optionally, estimation can be performed subject to inequality constraints, imposed by means of a logistic map. If upper and lower bounds UB and LB are specified, the reported parameter β_i is a logistic transformation of an underlying, unconstrained value ϕ_i ,

$$\beta_i = LB + (UB - LB)\frac{\exp(\phi_i)}{1+\exp(\phi_i)}. \tag{4.12}$$

Provided the constraint does not bind with $|\hat{\phi}_i| \to \infty$, approximate standard errors for $\hat{\beta}_i$ are computed by the delta method.

4.2.6 Polynomial Distributed Lags

The Data Editing and Transformation dialog contains an option to create moving averages with the form

$$z_{it} = \frac{\sum_{j=0}^{N}(N-j+1)^i x_{t-j}}{\sum_{j=0}^{N}(N-j+1)^i}, \quad i = 0,\ldots,4 \tag{4.13}$$

where x_t is a variable in the data set and N is a chosen lag length. If all or some of these variables are included in an equation as Type 1 regressors, with coefficients α_i , this is equivalent to including lags $x_t,\ldots,x_{t-N}$ with coefficients

©James Davidson 2015

$$\beta_j = \sum_{i=0}^{4} \frac{(N-j+1)^i}{\sum_{j=0}^{N}(N-j+1)^i}\alpha_i, \quad j = 0,\ldots,N \tag{4.14}$$

These lie on a polynomial of order ≤ 4, and hence are constrained to vary smoothly. Note that suppressing the zero order term imposes the end-point constraint $\beta_{N+1} = 0$.

4.3 Bilinear Models

The *bilinear* option (see Priestley 1988, Tong 1990) replaces equation (4.1) with

$$\phi(L)w_t = \gamma_{02} + \boldsymbol{\pi}_2'\boldsymbol{x}_{2t} + \big(\lambda(L)v_{t-1}\big)\psi(L)w_{t-1} + \theta(L)v_t \tag{4.15}$$

where

$$w_t = (1-L)^{d_1}(Y_t - \gamma_{10} - \gamma_1 t - \boldsymbol{\pi}'\boldsymbol{x}_{1t}), \quad v_t = \boldsymbol{\pi}_3'\boldsymbol{x}_{3t} + u_t\,. \tag{4.16}$$

In (4.15), $\psi(L) = \psi_1 + \psi_2 L + \ldots + \psi_p L^{p-1}$ where p is also the order of the AR polynomial $\phi(L)$ and $\lambda(L) = 1 + \lambda_1 L + \ldots + \lambda_{r-1}L^{r-1}$, similarly. This is a restricted version of the BL(p, q, m, r) class of models specified by Subba Rao (1981). The models are equivalent in the case $p = m$ and $r = 1$, but for $r > 1$, our case is restricted to have just $p + r$ additional parameters relative to the ARMA(p, q) case, instead of pr additional parameters where the coefficients of $w_{t-j}v_{t-k}$, $j = 1,\ldots,p$, $k = 1,\ldots,r$ are unrestricted.

The bilinear model can also be implemented in combination with ECM terms and nonlinear features. Simply modify equations (4.21) or (4.29) with the new features of (4.15)+(4.16). Note that $p \neq m$ can be implemented by fixing parameters at zero.

In this model, note that including an intercept and/or trend term in x_{2t} has a different effect from using the built-in dummies, which play the same role in the dynamics as variables in x_{1t}. Intercept and trend dummies can be included in $\boldsymbol{x}_{2t}$ as ordinary regressors, but in this case don't forget to deselect the built-in intercept and trend.

4.4 Nonlinear Moving Average Models

Several schemes have been proposed recently for modelling processes which switch stochastically between stationary and nonstationary behaviour, depending on the size of shock. Large shocks tend to have permanent effects, while the effect of small shocks is transient. The basic scheme is of the form

$$\Delta Y_t = u_t - g_{t-1}u_{t-1} \tag{4.17}$$

where g_t is a function depending on u_t that interpolates between 0 and 1. For example, in the STOPBREAK model of Engle and Smith (1999),

$$g_t = 1 - \frac{u_t^2}{\gamma + u_t^2}, \quad \gamma > 0 \tag{4.18}$$

so that g_t varies inversely with the magnitude of the shock. In the STIMA model of Gonzalo and Martinez (2003),

$$g_t = \begin{cases} \theta_1, & |u_t| < r \\ \theta_2, & |u_t| \geq r. \end{cases} \tag{4.19}$$

©James Davidson 2015

where for direct comparability with the STOPBREAK model we would have $\theta_1 = 1$ and $\theta_2 = 0$.

TSM implements a form suggested by Gonzalez (2004), called the "smooth permanent surge" or SPS model, of which special cases can closely approximate (4.18) and (4.19). This can be written as

$$g_t = \alpha + \frac{\beta}{1 + \exp\{-\gamma(u_t - c_1)(u_t - c_2)\}}. \tag{4.20}$$

where $\gamma > 0$, and $c_2 > c_1$ by convention. This depends on five parameters that can all be freely estimated, in principle, although fixing certain of them at given values yields the special cases indicated. Thus, setting $\alpha = 2$, $\beta = -2$ and $c_1 = c_2 = 0$ yields a function depending on $\gamma > 0$, which in common with the STOPBREAK function lies close to 1 when u_t^2 is small, and smoothly approaches 0 as u_t^2 increases. On the other hand, setting $c_1 = -r$ and $c_2 = r$, and $\gamma = 100$ (or any sufficiently large value) gives a close approximation to (4.19) with $\theta_1 = \alpha$ and $\theta_2 = \alpha + \beta$.

4.5 Error Correction Models

Equation (4.1) can be modified as

$$\phi(L)(1-L)^{d_1}(Y_t - \gamma_{01} - \gamma_1 t - \boldsymbol{\pi}_1'\boldsymbol{x}_{1t}) = \gamma_{02} + \boldsymbol{\pi}_2'\boldsymbol{x}_{2t} + \boldsymbol{\upsilon}'\boldsymbol{Z}_{t-K} + \theta(L)v_t \tag{4.21}$$

where $\boldsymbol{Z}_t$ is a $S \times 1$ vector of *equilibrium relations*, and $\boldsymbol{\upsilon}$ is a $S \times 1$ vector of *loadings coefficients*. The lag $K \geq 1$ is selectable, While in a single equation $S = 1$ would be the typical case, the only restrictions needing to be observed by the nonlinear mapping is to ensure parameters are identified – in general, no more unrestricted parameters than independent variables.

4.5.1 Equilibrium Relations

Two schemes are implemented to form the elements of $\boldsymbol{Z}_t$ as linear combinations of specified variables.

1. Set

$$\boldsymbol{Z}(\boldsymbol{x}_{7t}) \;=\; \boldsymbol{\Pi}_7 \boldsymbol{x}_{7t} \tag{4.22}$$

   where $\boldsymbol{\Pi}_7$ is a $P \times S$ matrix of coefficients of S equilibrium relations in P variables, of which one element of each column must be normalized to 1. When the data are nonstationary and a unit root is imposed, these relations are commonly called *cointegrating*, although such a model is also compatible with stationary data. The matrix must have sufficient restrictions imposed to identify its remaining elements. See Davidson (1998, 2000) for details. The vector $\boldsymbol{x}_{7\,t}$ may include the dependent variable.

2. Let $S = 1$ and define

$$Z_t = Y_t - \gamma_{01} - \gamma_1 t - \boldsymbol{\pi}_1'\boldsymbol{x}_{1t} \tag{4.23}$$

   where the parameters are constrained to match those in equation (4.1). This allows nonlinear autoregressions to be implemented, see the next section. To include regressors not subject to the implicit coefficient restrictions of (4.23), these can be included as Type 2.

©James Davidson 2015

4.5.2 Fractional Cointegration

In the usual case, the vector υ is constant with elements υ_i. However, it can optionally be replaced by a vector of lag polynomials $\upsilon(L)$ with elements $\upsilon_i(1-L)^{d_1-d_{3i}}$ for $i = 1,\ldots,S$ where d_1 is the fractional differencing coefficient in (4.21) and $0 \le d_{3i} \le d_1$. This is the *fractional cointegration* model. The usual "full cointegration" case is $d_{3i} = d_1$.

4.5.3 Nonlinear Error Correction and Nonlinear AR

A further variant of (4.21) is to replace the ECM term $\upsilon'\boldsymbol{Z}_{t-K}$ by $\upsilon' f_2(\boldsymbol{Z}_{t-K})$ where $f_2(.)$ represents a vector, of the same dimensions as its argument, whose elements are transformations of the corresponding elements of the argument. The usual linear case is $f(x) = x$. However, other programmed options available include

Exponential Smooth Transition: $f_2(x) = x[1 - \exp(-\zeta x^2)],\ \zeta > 0$ (4.24)

Asymmetric: $f_2(x) = x + \zeta|x|$ (4.25)

Cubic Polynomial: $f_2(x) = (x - \zeta)x^2$ (4.26)

where ζ is in each case an additional parameter. In these cases the pair of parameters (υ, ζ) determine the error-correcting behaviour, although be careful to note that in (4.24) each parameter is unidentified when the other is 0, whereas in (4.25) and (4.26), ζ is unidentified when $\upsilon = 0$. This makes significance tests problematic, although note that the parameters can be fixed at chosen values as well as estimated.

When option (4.23) is selected in conjunction with (4.24), (4.25) or (4.26), the model implements a nonlinear autoregressive specification. For example, (4.24) combined with a linear AR term corresponds to the ESTAR model. This model interpolates smoothly between two AR coefficients, depending on the path of the process. If a unit root is imposed and $\upsilon < 0$, it embodies "target zone"-type behaviour in which the process resembles a random walk for small deviations, but reverts to the mean after large deviations. An alternative way of implementing nonlinear dynamics is through the smooth-transition regime switching option. These models allow the value of exogenous variables to control the dynamic regime, but not the value of the ECM residual itself, as here. (Note: treatments of the ESTAR model sometimes allow a different lag on the two occurrences of x in (4.24). This option is not available in the present implementation.)

4.6 User Coded Functions

4.6.1 Coded Formulae

A virtually unlimited range of nonlinear specifications can be implemented by entering mathematical formulae directly. In other words, a formula can be typed using natural notation, combining the operators +, –, *, /, ^, parentheses, and standard mathematical functions (log, exp, sin, etc.) with parameters and variables. As a simple example, to compute the nonlinear regression

$$y_t = \alpha + \beta x_t^{\gamma} + u_t \tag{4.27}$$

where the data set contained corresponding variables WYE and EXE, one could simply enter the line

```
WYE = alpha + beta*EXE^gamma
```

See the User's Manual for detailed coding instructions.

©James Davidson 2015

Another possibility is to enter an implicit formula for the residual. For example, to estimate the Box-Cox (1964) transformed regression

$$\frac{y_t^\gamma - 1}{\gamma} = \alpha + \beta \frac{x_t^\gamma - 1}{\gamma} + u_t \tag{4.28}$$

one would select the "Residual" option and enter the line.

```
(WYE^gamma -1)/gamma - alpha - beta*(EXE^gamma - 1)/gamma
```

Note that in this case no "=" should appear, since the formula is an implicit representation of the residual. The formula to represent (4.27) could likewise be entered under this option as

```
WYE - alpha - beta*EXE^gamma
```

but the "Equation" style is to be preferred whenever it exists, since it identifies the normalized variable for the construction of simulations, forecasts etc..

4.6.2 Formula Types

Coded formulae can enter the model in five different ways, represented by the symbols f_1, f_2, f_3, and f_4 in

$$\begin{aligned}\phi(L)(1-L)^{d_1}(f_1(Y_t, \boldsymbol{x}_{4t}; \boldsymbol{\xi}) - \gamma_{01} - \gamma_1 t - \boldsymbol{\pi}_1' \boldsymbol{x}_{1t}) &= \gamma_{02} + \boldsymbol{\pi}_2' \boldsymbol{x}_{2t} \\ &+ f_2(\boldsymbol{x}_{4t}; \boldsymbol{\xi}) + \upsilon' f_3(\boldsymbol{Z}_{t-K}, \boldsymbol{x}_{4t}; \boldsymbol{\xi}) + \theta(L)(v_t - f_4(v_{t-j}, j > 0, \boldsymbol{x}_{4t}; \boldsymbol{\xi}))\end{aligned} \tag{4.29}$$

and also by $\boldsymbol{f}_5$ in

$$\boldsymbol{Z}_t = \boldsymbol{f}_5(\boldsymbol{x}_{7t}, \boldsymbol{\xi}) \quad (S \times 1). \tag{4.30}$$

However, note that a model can contain only <u>one</u> formula at a time. These five options are mutually exclusive, since they share the same locations to store specifications and parameter values. Thus, the symbols for the variables ($\boldsymbol{x}_{4t}$) and parameters ($\boldsymbol{\xi}$) are generic here, to be defined in context.

4.6.3 Coded Equations

As noted, there are two options for representing f_1, the usual case being the "Equation" form

$$f_{1t} = Y_t - g_1(f_{1,t-j}, j > 0, \boldsymbol{x}_{4t}; \boldsymbol{\xi}). \tag{4.31}$$

Thus, note how the equation can have a recursive form, allowing specialized forms of nonlinear dynamics. The symbol E#{-1} is used to represent the lagged residual in formulae. Since this option allows a normalized ("left-hand-side") variable to be specified, fitted values, simulations and forecasts can be generated. This is not possible for the implicit ("Residual") form

$$f_{1t} = g_1(Y_t, f_{1,t-j}, j > 0, \boldsymbol{x}_{4t}; \boldsymbol{\xi}) \tag{4.32}$$

which accordingly should only be used on cases such as (4.28), where a suitable normalization does not exist.

In the "Equation" coding style the disturbance is added automatically to the formula, and does not need to appear explicitly. For models to be used for stochastic simulation, another style of coding is available in which the disturbance appears explicitly. Thus, equation (4.27) might be coded as

```
WYE = alpha + beta*EXE^gamma + W#
```


©James Davidson 2015

where W# is a reserved name denoting the artificially generated disturbance. It can also appear lagged, as W#{j} for $j > 0$. The disturbance can be transformed or can enter the model nonlinearly, providing much greater modelling flexibility. However, this coding style cannot be used for estimation purposes, since such equations cannot be inverted to generate residuals. Attempting to estimate an equation containing W# produces an error. Instead, create a separate model with the "Residual" option. The reserved name E#{-j} can be used denote the residual lagged $j > 0$ periods in this case.

4.6.4 Coded Component

The option f_2 allows a nonlinear function of regressors $\boldsymbol{x}_{4t}$ to appear in the dynamic part of the model. The dependent variable(s) can appear in linear autoregressive form. Note that without an autoregressive or fractional component it would be equally possible to set up this nonlinear specification through f_1, but otherwise the dynamics will be different, similarly to the distinction between Type 1 and Type 2 regressors.

4.6.5 Coded Error Correction Mechanism.

Pre-programmed formulae for the option f_3 have already been defined in (4.24), (4.25) and (4.26). It will always be computationally more efficient to use these pre-programmed cases, so do not attempt to code them unless variations are desired. Also note that the pre-programmed cases can be combined with either f_1, f_2 or f_4.

4.6.6 Coded Moving Average Model

A fairly general pre-programmed form for f_4 is already defined in (4.20), depending on up to five parameters. Coding this function allows other variants to be implemented, including system variants. Note that the coded function must correspond to $g_{t-1}u_{t-1}$ in (4.17), which allows the lag to be modified as well as the functional form.

Note: recursive formulae require T calls to the parsing routine at each function evaluation instead of one, where T denotes sample size. Accordingly, these models are relatively more expensive computationally than static formulations.

4.6.7 Coded Equilibrium Relations

Equation (4.30) provides a third option alongside (4.22) and (4.23) to define equilibrium relations. These are entered in the same way as implicit formulae for the residuals. A natural application for this feature is to introduce structural shifts in coefficients, which are easily coded using dummy variables. Be careful to distinguish between nonlinear equilibrium conditions ($\boldsymbol{f}_5$) and nonlinear ECMs (f_4). These can coexist provided the latter is one of the pre-programmed options (4.24)-(4.26).

4.6.8 Ox Coding

Although there is no practical limit to the size of interactive formulae, they are still limited to a single line of code. An alternative option for coding models is to write an Ox function with the requisite code, and compile this with the program. Such functions can be arbitrarily complex, taking as many lines as necessary and employing loops, conditional statements, function calls, etc. Such code may also run significantly faster than interactive formulae of comparable complexity.

This option is available for creating model components of type f_1, which can of course represent complete models. There are also options to return a complete likelihood functions, and test statistics, either using estimation outputs or as free-standing functions of the data. By creating an explicit model solution separately, this approach can also be used

 ©James Davidson 2015

to create simulations and forecasts for models expressed in implicit form for estimation. Finally, Ox code can be supplied for the generation of artificial data for simulation exercises.

This approach requires some knowledge of programming basics and of the Ox language, although significantly less than would be required to create an Ox program from scratch. Moreover, all the usual features of the package including estimation, diagnostic testing, simulation and forecasting are available just as for the pre-programmed models.

Ox coding cannot be combined with interactive formulae nor, in the current release, there is any option to create functions of type f_2, f_3 or f_4 as Ox code. These features can of course be incorporated into a complete Ox-coded model.

4.7 Discrete Data Models

4.7.1 Probit and Logit Models

In probit and logit models for binary data, the probability of the binary dependent variable Y_t taking the value 1 is modelled as $F(z_t)$ where $F(.)$ denotes respectively the standard normal and logistic CDF, and

$$z_t = \gamma_{02} + \gamma_1 t + \boldsymbol{\pi}_1' \boldsymbol{x}_{1t} + \boldsymbol{\pi}_2' \boldsymbol{x}_{2t} + \upsilon(L)' f_2(\boldsymbol{Z}_{t-K}) \tag{4.33}$$

is a continuously distributed latent process. ECMs can be included in this equation, using either of the formulations (4.22) and (4.30) for the equilibrium condition. Endogenous dynamics are also possible, by including Y_{t-j} for $j > 0$ in the vector $\boldsymbol{x}_{2t}$.

Optionally, the chi-squared can be used for the distribution of the latent probit variable. In this case, the degrees of freedom of the distribution becomes an additional parameter to be estimated or (optionally) fixed.

4.7.2 Ordered Probit and Logit

These models allow the analysis of more than two discrete states, provided these are ordered monotonically as functions of the explanatory variables. Essentially, the same model must explain the probabilities of the states apart from shifts of intercept.

Suppose for example that Y_t can assume the values 0, 1, 2, or 3. These might correspond to responses varying from negative to positive in a sample survey. The probabilities of these states would be modelled as

$$\begin{aligned} P(Y_t = 0) &= F(-z_t) \\ P(Y_t = 1) &= F(\gamma_1 - z_t) - F(-z_t) \\ P(Y_t = 2) &= F(\gamma_1 + \gamma_2 - z_t) - F(\gamma_1 - z_t) \\ P(Y_t = 3) &= 1 - F(\gamma_1 + \gamma_2 - z_t), \end{aligned} \tag{4.34}$$

where F denotes the normal or logistic CDF as before, z_t is defined by (4.33) as before, and γ_1 and γ_2 are two additional parameters to be estimated.

Note that the γ_j parameters are constrained to be non-negative, and are necessarily positive unless the corresponding category is empty. In the latter case the corresponding term is omitted from the likelihood function, in effect, and the parameter is unidentified. This problem can be overcome in estimation by fixing the parameter in question at 0.

©James Davidson 2015

4.7.3 Count Data

In count data models the data are again integer-valued, but in this case there is no upper bound the set of possible values.

In the Poisson model, the probabilities that $Y_t = 0, 1, 2, 3, \ldots$ are modelled using the Poisson distribution with conditional mean $\phi_t = E(Y_t|x_{1t}, x_{2t}, x_{3t})$ where optionally either

$$\phi_t = z_t \tag{4.35}$$

or

$$\phi_t = \exp(z_t). \tag{4.36}$$

where z_t is defined by (4.33). In the negative binomial models, the variance of Y_t has the representation

$$\mathrm{Var}(Y_t|x_{1t}, x_{2t}, x_{3t}) = \phi_t\,(1+ \alpha\phi_t^{1-k}) \tag{4.37}$$

where $k = 1$ in the NegBin I case and $k = 0$ in the NegBin II case. The Poisson is the case $\alpha = 0$. See Cameron and Trivedi (1986) for details of these cases.

4.7.4 Autoregressive discrete models

A further dynamic variant is to replace z_t in (4.33) by $\lambda(L)z_t$ where

$$\lambda(L) = 1 - \lambda_1 L - \cdots - \lambda_p L^p\,. \tag{4.38}$$

As an example, consider the Poisson specification (4.35) and

$$z_t = \gamma_{02} + \pi_{21} Y_{t-1} + \pi_{22} z_{t-1} \tag{4.39}$$

where all the parameters are positive. It is known that the stochastic process $\{Y_t\}$ is stationary and ergodic if $\pi_{21} + \pi_{22} < 1$ (see Fokianos et al., 2009). The lags $z_{t-1}, \ldots, z_{t-p}$ are specified formally in the software as Type 2 regressors, so that this structure is distinct from the ARFIMA setup in (4.1), which is of course not available in the discrete case. Therefore, note that the intercept is only available in the Type 2 form, in this specification.

4.7.5 Zero-inflated Poisson and ordered Probit

In some data sets, the number of zero cases may appear excessive relative to the assumed distribution of the cases. To deal with this phenomenon, it may be hypothesized that the observations are drawn from two regimes, one of which yields a zero, the other a Poisson or ordered Probit drawing. The probability of the drawing coming from the first regime can be modelled by a separate regression function, having the general form $F(w_t)$ where

$$w_t = \gamma_{03} + \boldsymbol{\pi}_3' \boldsymbol{x}_{3t} \tag{4.40}$$

and F denotes the Gaussian CDF in the ordered Probit case, and the logistic in the Poisson and negative binomial cases.

©James Davidson 2015

5 Systems of Equations

5.1 The Basic Model

Now let Y_t denote a $N \times 1$ vector of jointly determined variables. The generalization of equation (4.21) takes the form

$$\begin{aligned}\Phi(L)\Delta^{d_1}(\mathbf{B}\boldsymbol{Y}_t - \boldsymbol{\gamma}_{01} - \boldsymbol{\gamma}_1 t - \boldsymbol{\Pi}_1 \boldsymbol{x}_{1t}) \\ = \boldsymbol{\gamma}_{02} + \boldsymbol{\Pi}_2 \boldsymbol{x}_{2t} + \Upsilon(L) f_2(\boldsymbol{Z}_{t-k}) + \Theta(L)(\boldsymbol{\Pi}_3 \boldsymbol{x}_{3t} + \boldsymbol{u}_t)\end{aligned} \tag{5.1}$$

Note that multiple equations are not available for bilinear or discrete data models.

5.2 Definitions and Details

5.2.1 System Notation

In equation (5.1),

$$\Phi(L) = \boldsymbol{I} + \Phi L + \ldots + \Phi_p L^p \tag{5.2}$$

and

$$\Theta(L) = \boldsymbol{I} + \Theta L + \ldots + \Theta_q L^q \tag{5.3}$$

are $N \times N$ matrices of lag polynomials. Δ^{d_1} is a $N \times N$ diagonal matrix with elements $(1-L)^{d_{1j}}$ on the diagonal for $j = 1,\ldots,N$. Be careful to note that in the VAR and VMA models, lags on all the endogenous variables appear in each equation, by default. To fit a restricted model with (e.g.) only own-lags included, restrictions can be imposed individually in the Values dialogs.

5.2.2 System Exogenous Variables

Constant coefficient matrices are $\boldsymbol{\gamma}_{01}$, $\boldsymbol{\gamma}_{21}$ and $\boldsymbol{\gamma}_1$, $N \times 1$, and $\boldsymbol{\Pi}_1$, $\boldsymbol{\Pi}_2$, $\boldsymbol{\Pi}_3$ are matrices with N rows, conformable with $\boldsymbol{x}_{1t}$, $\boldsymbol{x}_{2t}$, $\boldsymbol{x}_{3t}$ respectively. Note that by default, the specification of every equation is the same. Individual zero-restrictions can be imposed on an individual basis in the Values dialogs.

5.2.3 Simultaneous Equations

To specify that an equation contains current endogenous variables, such that the matrix $\boldsymbol{B}$ in equation (5.1) is non-diagonal, these variables must be included as regressors of Type 1 in the equation (that is, as elements of $\boldsymbol{x}_{1t}$). When one or more endogenous variables are present in this set, the program automatically computes the FIML estimator (see Estimation Criteria below). In this case the corresponding elements of $\boldsymbol{\Pi}_1$ are either to be interpreted as non-diagonal elements of $\boldsymbol{B}$ or, in the case of the normalized (left-hand-side) variable of the equation, are automatically fixed at 0**.** Some of these coefficients must be subject to identifying restrictions to ensure consistent estimation. The user *must* use the Values / Equation dialog to impose these restrictions on each equation, typically by fixing some coefficients to 0.

5.2.4 Nonlinear Systems

In equation (5.1) the vector $\boldsymbol{Y}_t$ can be replaced by a vector $\boldsymbol{f}_1(\boldsymbol{Y}_t, \boldsymbol{x}_{4t}; \boldsymbol{\xi})$ where

©James Davidson 2015

$$\boldsymbol{f}_{1t} = \boldsymbol{Y}_t - \boldsymbol{g}_1(\boldsymbol{f}_{1,t-j}, j > 0, \boldsymbol{x}_{4t}; \boldsymbol{\xi}). \tag{5.4}$$

Exactly as for the single-equation case, these functional forms can be either coded with the built-in formula parser, or supplied as an Ox function. It's important to note that these functions *cannot be simultaneous*; in other words, $\boldsymbol{x}_{4t}$ may *only* contain predetermined (exogenous or lagged endogenous) variables.[1] However, *linear* simultaneity is permitted, in the sense that $\boldsymbol{B}$ is allowed to be non-diagonal in (5.1). If elements of $\boldsymbol{Y}_t$ are included in $\boldsymbol{x}_{1t}$, the corresponding coefficients of $\boldsymbol{\Pi}_1$ are treated as non-diagonal elements of $\boldsymbol{B}$, or fixed at 0, just as in the linear model case (see preceding paragraph). Be careful to note that these are treated as coefficients of $\boldsymbol{f}_{1t}$, not $\boldsymbol{Y}_t$!

The implicit nonlinear specification

$$\boldsymbol{f}_{1t} = \boldsymbol{g}_1(\boldsymbol{Y}_t, \boldsymbol{f}_{1,t-j}, j > 0, \boldsymbol{x}_{4t}; \boldsymbol{\xi}) \tag{5.5}$$

is also permitted, again subject to the restriction that each element of $\boldsymbol{f}_{1t}$ is a function of one and only one element of $\boldsymbol{Y}_t$. The warning in footnote 1 applies here too. In addition, please note that models of this type *cannot be simulated*. Simulations will be computed, but will correspond to an incorrect model. Instead, create a separate model for simulation. This can optionally use the reserved names W#1, W#2, etc., to represent the disturbances on each equation.

5.3 Error Correction Models

Two types of vector error-correction model (VECM) are implemented, 'open loop' and 'closed loop'. In open loop models, the equilibrium relations are specified as in (4.22) or (4.30) with $\boldsymbol{x}_{7t}$ a $P\times1$ vector that may include any specified endogenous variables (elements of $\boldsymbol{Y}_t$) as well as exogenous variables.

Closed-loop models are implemented by generalizing the representation in (4.23). In other words, let

$$\boldsymbol{Z}_t = \boldsymbol{\Pi}_7(L)\,(\mathbf{B}\boldsymbol{Y}_t - \boldsymbol{\gamma}_{01} - \boldsymbol{\gamma}_1 t - \boldsymbol{\Pi}_1 \boldsymbol{x}_{1t}). \tag{5.6}$$

noting that the matrix $\boldsymbol{\Pi}_7(L)$ of cointegrating coefficients can optionally include lag polynomial factors.

In the standard (open loop or closed loop) VECM model, the $N \times S$ matrix $\boldsymbol{\Upsilon}$ is a matrix of constant loadings coefficients υ_{ji}, weighting the ith lagged equilibrium relation in equation j of the system. Optionally, it can be replaced by the matrix $\boldsymbol{\Upsilon}(L)$ containing elements $\upsilon_{ji}(1-L)^{d_{1j}-d_{3ji}}$ where d_{1j} is defined **5.2.1** above, and the d_{3ji} are additional parameters. This is the fractional VECM (FVECM) model, and this case is called *regular fractional cointegration*. The equilibrium relations are potentially cointegrating in the sense that they are integrated to order $d_{1j} - d_{3ji} < d_{1j}$. While this value must be the same for all j, d_{1j} and d_{3ji} can depend on j. d_{3ji} can potentially differ with respect to i when variables with different integration order appear in different cointegrating relations. The standard VECM model corresponds to the case $d_{3ji} = d_{1j} = 1$, all j and i.

[1] CAUTION: it is the user's responsibility to respect this requirement! The program cannot monitor a nonlinear specification. If this violates the exogeneity restriction, the estimates will not correspond to a valid estimator, and will be inconsistent in general. Simulations will also be incorrect.

©James Davidson 2015

In the closed-loop model, in which $P = N$ and (5.6) represents the equilibrium relations, a constant loadings matrix $\boldsymbol{\Upsilon}$ can be optionally combined with a cointegrating matrix $\boldsymbol{\Pi}_7(L)$ whose typical elements are of the form the form $\pi_{7,ji}(1-L)^{d_{1j}-d_{3ji}}$. Here, the jth element of the vector

$$\mathbf{B}\boldsymbol{Y}_{\mathrm{t}} - \boldsymbol{\gamma}_{01} - \boldsymbol{\gamma}_1 t - \boldsymbol{\Pi}_1 \boldsymbol{x}_{1t} \tag{5.7}$$

enters the ith cointegrating relation in fractionally differenced form with differencing parameter $d_{1j} - d_{3ji}$. This case of the FVECM is called *generalized fractional cointegration*.

Note that the generalized and regular cointegrating models are not distinguished when $d_{1j} - d_{3ji}$ does not depend on i or j. Also, be careful to note that the parameter d_{3ji} is unidentified if the jth variable is suppressed in the ith cointegrating relation. In this case it should be fixed to 0. Tests of zero restrictions on the ECM coefficients are also problematic, because the corresponding d_3 coefficients are unidentified under the null hypothesis. For more information on fractional cointegration models see Davidson (2002, 2005), and also Davidson, Byers and Peel (2006) for an example using TSM4.

In both models, the program allows the d_{1j} and d_{3ji} parameters for $j = 2,\ldots,N$ to be estimated as the differences from d_{11} and d_{31i}. This allows the natural restriction that all variables are integrated/cointegrated to the same order to be easily imposed and tested.

©James Davidson 2015

6 Conditional Heteroscedasticity Models

6.1 Single Equation Models

In any of (4.1)+(4.2) or (4.21)+(4.2) or (4.29)+(4.2) or (4.15)+(4.16), let $u_t = h_t^{1/2} e_t$ where $e_t \sim iid(0,1)$. h_t can be defined either by

$$\begin{aligned}\beta(L)(h_t^{\eta/2} - \omega - \boldsymbol{\pi}_4'\boldsymbol{x}_{4t}) \\ = \kappa + \boldsymbol{\pi}_5'\boldsymbol{x}_{5t} + [\beta(L) - \delta(L)(1 + \alpha((1-L)^{d_2} - 1))]((1 + \mu s_t)\,|\,u_t\,|^{\eta} + \boldsymbol{\pi}_6'\boldsymbol{x}_{6t})\end{aligned} \tag{6.1}$$

where $s_t = I(u_t < 0)$ (the GARCH class of models with $\eta = 2$, or the APARCH class with $\eta > 0$ unrestricted) or by

$$\begin{aligned}\beta(L)(\log h_t - \omega - \boldsymbol{\pi}_4'\boldsymbol{x}_{4t}) \\ = \kappa + \boldsymbol{\pi}_5'\boldsymbol{x}_{5t} + [\beta(L) - \delta(L)(1 + \alpha((1-L)^{d_2} - 1))](g_t + \boldsymbol{\pi}_6'\boldsymbol{x}_{6t})\end{aligned} \tag{6.2}$$

where

$$g_t = h_t^{-1/2}\,|\,u_t\,| - m + \tau h_t^{-1/2} u_t \tag{6.3}$$

(the EGARCH class). Only one of ω and κ in (6.1) and (6.2) may be different from zero.

Here, $\delta(L)$ and $\beta(L)$ are finite lag polynomials analogous to $\phi(L)$ and $\theta(L)$ in (4.1) The $\boldsymbol{x}_{jt}$ for $j = 4,5,6$ are vectors of variables with coefficient vectors $\boldsymbol{\pi}_j$. Also note that the vectors $\boldsymbol{x}_{1t}$, $\boldsymbol{x}_{2t}$ or $\boldsymbol{x}_{3t}$, in equation (4.1) or (4.15)/(4.16), can include either h_t or $h_t^{1/2}$. This is the GARCH-M model. It might often be appropriate to treat the GARCH-M variable as a Type 3 regressor. It cannot be of Type 1 when the data are differenced, as in (4.6).

The basic ARCH and GARCH models are implemented as the special case of (6.1). In the usual Bollerslev (1986) notation, the GARCH model is written as

$$\beta(L)h_t = \kappa + \alpha(L)u_t^2 \tag{6.4}$$

where $\alpha(L) = \beta(L) - \delta(L)$ and $\kappa = \beta(1)\omega$.

6.2 Definitions and Details

6.2.1 GARCH Parameterization

In equations (6.1) and (6.2), observe that the zero-order term of the lag polynomial on the right-hand side is zero by construction, so that the model only involves lagged values of u_t^2. The usual GARCH formulation is obtained if only the parameters ω, $\beta(L)$ and $\delta(L)$ are present. The parameterisation adopted in (6.1) is the 'ARMA-in-squares' form

$$\delta(L)u_t^2 = \kappa + \beta(L)v_t \tag{6.5}$$

where $v_t = u_t^2 - h_t$ and $\kappa = \beta(1)\omega$. The roots of the polynomial $\delta(L)$ must be stable for covariance stationarity. The package will optionally report the estimates in the conventional (Bollerslev 1986) form, as in equation (6.4). As a simple example, the GARCH(1,1) model may be written either as

$$h_t = \kappa + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} \tag{6.6}$$

or as

©James Davidson 2015

$$u_t^2 = \kappa + \delta_1 u_{t-1}^2 + v_t - \beta_1 v_{t-1} \tag{6.7}$$

where $v_t = x_t^2 - h_t$, $\delta_1 = \alpha_1 + \beta_1$. The program can report either the (α_1, β_1) pair, or the $(\delta_1, -\beta_1)$ pair. Whether κ or $\omega = \kappa/\beta(1)$ is estimated is also a user-selectable option. We refer to κ as the "GARCH intercept of Type 2" by analogy with equation (4.1) – see Section 4.2.3 for the details.

The lag structures in (6.1) and (6.2) are the same, for convenience and comparability. This can be thought of as a flexible dynamic form, with general application, but note that the interpretation of restrictions is different in each case. The case $\alpha \geq 1$ in (6.2) has no implications for covariance stationarity, as it does in equation (6.1). Abstracting from the role of possible exogenous variables, the EGARCH model is stationary if the lag coefficients in (6.2) are square summable (see Nelson 1991). This is satisfied here provided the roots of $\beta(L)$ are stable and $d > -0.5$.

6.2.2 HYGARCH and FIGARCH

The HYGARCH model (Davidson 2004a) specifies α in (6.1), referred to as the amplitude parameter. The case $\alpha = 1$ is the FIGARCH model, with d_2 the hyperbolic memory parameter. Note that setting $0 < \alpha < 1$ gives a stationary process, while with $\alpha \geq 1$ (which includes FIGARCH) it is nonstationary. If $d_2 = 1$, α reduces to the status of an additional autoregressive root, and so gives the IGARCH model with $\alpha = 1$.

6.2.3 Asymmetric GARCH and Power GARCH

In equation (6.1), $s_t = 1$ if $u_t < 0$ and 0 otherwise. μ is the so-called 'leverage' (asymmetry) parameter, which permits positive and negative disturbances to contribute differently to the conditional variance. If $\mu > 0$, there is a larger contribution when $u_t < 0$ than otherwise. In the case where η is fixed at 2, this supplies a variant of the 'threshold GARCH' or 'GJR' model (Glosten et al. 1993). In the case where $\eta > 0$ is a free parameter, (6.1) is equivalent to the 'asymmetric power ARCH' (APARCH) model (Ding et. al. 1993). Note that all these models can be extended with the natural FIGARCH/HYGARCH generalizations.

6.2.4 GARCH Regressors

The variables $\boldsymbol{x}_{4t}$, $\boldsymbol{x}_{5t}$ and $\boldsymbol{x}_{6t}$ will be called GARCH regressors of Types 1–3, in parallel with the exogenous components of the mean equation. In (6.1), these variables ought to be nonnegative to ensure the model always yields a positive value for the conditional variance. There is a program option to convert the selected regressors to absolute form automatically. The intercept parameter ω in (6.1) corresponds to the variance when all conditional heteroscedasticity effects are absent. Note different orders of lag of each Type can be specified, just as for regressors in mean.

6.2.5 EGARCH

The appearance of h_t on both sides of (6.2) is purely formal; since the zero-order lag term is zero in (6.2), as in (6.1), the equation is recursive. The parameter τ represents the degree of asymmetry effects, analogous to μ in equation (6.1), fixed at zero unless this model feature is selected. However, take care to note that the sign of has the reverse interpretation of that of μ. Note that the dynamic parameterization adopted precludes a separate parameter for the absolute shock. m is an optional term representing $E(h_t^{-1/2} \mid u_t \mid)$ according to the likelihood model selected. It can be replaced by zero, in which case the mean of g_t is absorbed by the intercept parameter, represented by either κ or $\beta(1)\omega$. By default, it is set as follows:

©James Davidson 2015

$m = \sqrt{2/\pi}$ if the likelihood is Gaussian;

$m = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)(\nu-1)}\sqrt{\frac{\nu}{\pi}}$ if the likelihood is Student's *t* with degrees of freedom ν;[2]

$m = \frac{\Gamma(2/\nu)}{\sqrt{\Gamma(3/\nu)\Gamma(1/\nu)}}$ if the likelihood is GED with parameter ν.[3]

6.2.6 HYEGARCH and FIEGARCH

By setting $\alpha \neq 0$ and $d_2 \neq 0$ in (6.2), hyperbolic decay models can be set up. Take care to note that the interpretation of the parameters, different from the case of (6.1). With $d_2 > 0$ the case of hyperbolic memory decay with summable hyperbolic coefficients is represented, with sum $1-(1-\alpha)\delta(1)/\beta(1)$. Call this the HYEGARCH model, as with HYGARCH the rate of decay varies inversely with d_2. The FIEGARCH case is where $d_2 < 0$, and here it is necessary for $\alpha < 0$ also, otherwise the hyperbolic lag coefficients have the wrong sign. In this case, the lag coefficients, and hence the autocovariances of $\log h_t$ are non-summable, and the process exhibits true long memory in volatility.

6.3 Conditional Heteroscedasticity in Discrete Data Models

In the probit and logit models the probability of the dependent variable Y_t taking the value 1 can be modelled as $F(h_t^{-1/2}z_t)$ where as before $F(.)$ denotes respectively the standard normal and logistic CDF, and either

$$h_t^{\eta/2} = 1 + \boldsymbol{\pi}_4'\boldsymbol{x}_{4t} + \boldsymbol{\pi}_5'\boldsymbol{x}_{5t} + \boldsymbol{\pi}_6'\boldsymbol{x}_{6t} \tag{6.8}$$

which we call the "GARCH/APARCH" model by analogy with equation (6.1), or

$$\log h_t = 1 + \boldsymbol{\pi}_4'\boldsymbol{x}_{4t} + \boldsymbol{\pi}_5'\boldsymbol{x}_{5t} + h_t^{-1/2}\boldsymbol{\pi}_6'\boldsymbol{x}_{6t} \tag{6.9}$$

which we call the "EGARCH" model by analogy with equation (6.2). The different explanatory variable Types in (6.8) and (6.9) exist mainly by default, as specializations of the usual models, but could be used to set up different lag structures.

The ordered probit and logit models are generalized in just the same way, with $P(Y_t = 0) = F(-h_t^{-1/2}z_t)$, $P(Y_t = j) = F(\gamma_1 + \cdots + \gamma_j - h_t^{-1/2}z_t) - F(\gamma_1 + \cdots + \gamma_{j-1} - h_t^{-1/2}z_t)$ for $j = 1,\ldots,J-1$ and $P(Y_t = J) = 1 - F(\gamma_1 + \cdots + \gamma_{J-1} - h_t^{-1/2}z_t)$.

In Poisson models, equations (4.35) and (4.37) can be modified as

$$E(Y_t|x_{1t}, \ldots, x_{7t}) = \phi_t \tag{6.10}$$

$$\mathrm{Var}(Y_t|x_{1t}, \ldots, x_{7t}) = \phi_t\,(1 + \alpha\phi_t^{1-k}) \tag{6.11}$$

where $\phi_t = \exp\{ h_t^{-1/2}z_t \}$, and h_t is given by (6.8) or (6.9).

---

[2] See Psarakis and Panetaros (1990)

[3] See Nelson (1991)

©James Davidson 2015

7 Conditionally Heteroscedastic Systems

7.1 Implemented Model Variants

In a system of equation with the form (5.1), the default generalizations of equations (6.1) and (6.2) are

$$\begin{aligned}&\mathbf{B}(L)(\boldsymbol{h}_t^{\eta/2}-\boldsymbol{\omega}-\boldsymbol{\Pi}_4\boldsymbol{x}_{4t})\\&\quad=\boldsymbol{\kappa}+\boldsymbol{\Pi}_5\boldsymbol{x}_{5t}+[\mathbf{B}(L)-(\boldsymbol{I}+\mathbf{A}(\Delta^{d_2}-\boldsymbol{I}))\boldsymbol{D}(L)]((\boldsymbol{I}+\hat{\boldsymbol{M}}\boldsymbol{s}_t)\,|\,\boldsymbol{u}_t\,|^{\eta}+\boldsymbol{\Pi}_6\boldsymbol{x}_{6t})\end{aligned}\tag{7.1}$$

where $\boldsymbol{s}_t=I(\boldsymbol{u}_t<0)$ and $\hat{\boldsymbol{M}}$ is a diagonal matrix of parameters, and

$$\begin{aligned}&\mathbf{B}(L)(\log\boldsymbol{h}_t-\boldsymbol{\omega}-\boldsymbol{\Pi}_4\boldsymbol{x}_{4t})\\&\quad=\boldsymbol{\kappa}+\boldsymbol{\Pi}_5\boldsymbol{x}_{5t}+[\mathbf{B}(L)-\boldsymbol{D}(L)(\boldsymbol{I}+\mathbf{A}(\Delta^{d_2}-\boldsymbol{I}))](\boldsymbol{g}_t+\boldsymbol{\Pi}_6\boldsymbol{x}_{6t})\end{aligned}\tag{7.2}$$

where

$$\boldsymbol{g}_t=\hat{\boldsymbol{H}}_t^{-1/2}(|\,\boldsymbol{u}_t\,|+\hat{\boldsymbol{T}}\boldsymbol{u}_t)-m\boldsymbol{\iota}\,.\tag{7.3}$$

Here, $\boldsymbol{u}_t=\hat{\boldsymbol{H}}_t^{1/2}\boldsymbol{e}_t$ ($N\times1$), where $\hat{\boldsymbol{H}}_t=\mathrm{diag}(\boldsymbol{h}_t)$ ($N\times N$), $\boldsymbol{e}_t\sim iid(\mathbf{0},\boldsymbol{C})$ ($N\times1$) and $\boldsymbol{C}$ ($N\times N$) is a fixed correlation matrix with units on the diagonal. $\hat{\boldsymbol{M}}$ and $\hat{\boldsymbol{T}}$ are diagonal matrices with equation asymmetry parameters on the diagonal, m is defined in **6.2.5** and $\boldsymbol{\iota}$ ($N\times1$) is the unit vector. Only one of the N-vectors $\boldsymbol{\omega}$ and $\boldsymbol{\kappa}$ in (7.1) and (7.2) may be different from zero. Refer to the discussion of the various special cases in Section **6.2**, which generalize in the natural way to the multivariate model.

The fixed-correlation restriction can be relaxed in the *dynamic conditional correlation* (DCC) model. Here it is assumed that $\boldsymbol{e}_t\sim iid(\mathbf{0},\boldsymbol{R}_t)$ where

$$\boldsymbol{R}_t=\mathrm{diag}(\boldsymbol{Q}_t)^{-1/2}\boldsymbol{Q}_t\mathrm{diag}(\boldsymbol{Q}_t)^{-1/2}\tag{7.4}$$

and

$$\boldsymbol{Q}_t=(1-\alpha-\beta)\boldsymbol{C}+\alpha\boldsymbol{e}_{t-1}\boldsymbol{e}_{t-1}'+\beta\boldsymbol{Q}_{t-1}\,.\tag{7.5}$$

In this implementation $\alpha>0$ and $\beta>0$ are additional scalar parameters, so that for $N>2$, all dynamic correlations follow the same process.

Alternatively, in the BEKK model (Engle and Kroner 1995), $\boldsymbol{u}_t=\boldsymbol{H}_t^{1/2}\boldsymbol{e}_t$ where $\boldsymbol{e}_t\sim iid(\mathbf{0},\boldsymbol{I})$ ($N\times1$) and

$$\begin{aligned}&\mathbf{B}^*(L)\mathrm{vec}(\boldsymbol{H}_t-\hat{\boldsymbol{\omega}}\boldsymbol{E}\hat{\boldsymbol{\omega}}-\boldsymbol{\Pi}_4\boldsymbol{x}_{4t}\boldsymbol{x}_{4t}'\boldsymbol{\Pi}_4')\\&\quad=\mathrm{vec}(\hat{\boldsymbol{\kappa}}\boldsymbol{C}\hat{\boldsymbol{\kappa}}+\boldsymbol{\Pi}_5\boldsymbol{x}_{5t}\boldsymbol{x}_{5t}'\boldsymbol{\Pi}_5')+\mathbf{A}^*(L)\mathrm{vec}(\boldsymbol{u}_t\boldsymbol{u}_t'+\boldsymbol{\Pi}_6\boldsymbol{x}_{6t}\boldsymbol{x}_{6t}'\boldsymbol{\Pi}_6')\end{aligned}\tag{7.6}$$

where

$$\mathbf{A}^*(L)=\mathbf{A}_1\otimes\mathbf{A}_1L+\cdots+\mathbf{A}_p\otimes\mathbf{A}_pL^p\quad(N^2\times N^2)\tag{7.7}$$

$$\mathbf{B}^*(L)=\boldsymbol{I}_{N^2}-\mathbf{B}_1\otimes\mathbf{B}_1L-\cdots-\mathbf{B}_q\otimes\mathbf{B}_qL^p\quad(N^2\times N^2)\tag{7.8}$$

Only one of $\hat{\boldsymbol{\omega}}=\mathrm{diag}(\boldsymbol{\omega})$ and $\hat{\boldsymbol{\kappa}}=\mathrm{diag}(\boldsymbol{\kappa})$ may be different from zero, and $\boldsymbol{C}$ and $\boldsymbol{E}$ are alternative correlation matrices.

©James Davidson 2015

7.2 Definitions and Details

7.2.1 Multivariate GARCH

In equations (7.1) and (7.2), **B**(L) and ***D***(L) are $N\times N$ lag polynomial matrices with zero-order terms equal to I_N. **A** is a $N\times N$ diagonal coefficient matrix, and $\boldsymbol{\Delta}^{d_2}$ is a $N\times N$ diagonal matrix with terms of the form $(1-L)^{d_{2i}}$ on the diagonal.

7.2.2 DCC Multivariate GARCH

This model is proposed by Engle (2002). It is available in either the standard or the "ARMA in squares" parameterizations, that is, it is possible to estimate either the pairs (α, β) or (δ, β) where $\delta = \alpha + \beta$. In the latter case it is easy to impose the stationarity condition $\delta < 1$. This choice is controlled by the same option as for the GARCH specification proper.

7.2.3 BEKK Multivariate GARCH

This model is suggested by Engle and Kroner (1995). Both diagonal and off-diagonal elements of the conditional covariance matrices are modelled, using a linear vector-ARMA structure to explain the evolution of vec($\boldsymbol{H}_t$) ($N^2 \times 1$). Considering for simplicity the case $p = q = 1$, note that this model is equivalent to

$$\begin{aligned}\boldsymbol{H}_t = {} & \hat{\boldsymbol{\kappa}}\boldsymbol{C}\hat{\boldsymbol{\kappa}} + \boldsymbol{\Pi}_5\boldsymbol{x}_{5t}\boldsymbol{x}_{5t}'\boldsymbol{\Pi}_5' + \mathbf{A}(\boldsymbol{u}_{t-1}\boldsymbol{u}_{t-1}' + \boldsymbol{\Pi}_6\boldsymbol{x}_{6t-1}\boldsymbol{x}_{6t-1}'\boldsymbol{\Pi}_6')\mathbf{A}' \\ & + \mathbf{B}\boldsymbol{H}_{t-1}\mathbf{B}' + (\boldsymbol{\Pi}_4\boldsymbol{x}_{4t}\boldsymbol{x}_{4t}'\boldsymbol{\Pi}_4' - \mathbf{B}\boldsymbol{\Pi}_4\boldsymbol{x}_{4t-1}\boldsymbol{x}_{4t-1}'\boldsymbol{\Pi}_4'\mathbf{B}')\end{aligned} \tag{7.9}$$

In the lag polynomial matrices in (7.7) and (7.8), the $\mathbf{A}_j$ and $\mathbf{B}_j$ matrices are specified in the Values dialogs and in the output, just like the corresponding matrices in (7.1) and (7.2). In other words, each row of these matrices is assigned for reporting purposes (arbitrarily, in this case) to the corresponding equation of the system. Be careful to interpret the results correctly, for the interpretation of these matrices is of course entirely different in the two cases.

Note three further points about the BEKK model.

a) The case implemented is the case $K = 1$ defined by Engle and Kroner (1995). This model conveniently has the same number of AR-MA parameters as the models in (7.1) and (7.2). The cases $K > 1$ are not implemented in this release.

b) The parameterization of the intercept matrix, as $\hat{\boldsymbol{\omega}}\boldsymbol{C}\hat{\boldsymbol{\omega}}$ where $\boldsymbol{C}$ is symmetric with unit diagonal and off-diagonals constrained inside [−1,1], exploits the common structure of the different models, and is also a convenient way to impose positive definiteness. Note $\hat{\boldsymbol{\kappa}}\boldsymbol{C}\hat{\boldsymbol{\kappa}} = \mathbf{B}\hat{\boldsymbol{\omega}}\boldsymbol{E}\hat{\boldsymbol{\omega}}\mathbf{B}'$ represents an alternative parameterization.

c) Be careful to note that the parameterization represented by (7.7) and (7.8) is not unique. For example, if all nonzero elements of **B** or **A** are negative, this gives the same likelihood as when all are positive, so the likelihoods are liable to be multimodal. This implies global under-identification, although the model will be locally identified in general. This fact has no practical importance provided the results are interpreted correctly.

©James Davidson 2015

8 Regime Switching

The switching options allow the model driving the series to switch stochastically between alternative cases of the conditional mean and variance equations. For example, a process might behave as a random walk in one regime, and as a stable process in another. The models can be written in generic form as

$$u_t(\theta(S_t)) = h_t(\theta(S_t))^{1/2} e_t \tag{8.1}$$

where $S_t = 1,\ldots, M$ is a dummy variable denoting the regime prevailing a time t, and $\theta(S_t)$ represents the parameter values applying in regime S_t. Here, the vector θ is to be thought of as the concatenation of all the parameters in the specified model. The vectors $\theta(1),\ldots, \theta(M)$ are estimated, although the package allows different groups of parameters to either switch, or be constrained equal across regimes. The current version of the program allows $M = 2$, 3 or 4.

Three types of switching mechanism are supported in the package. See Kim and Nelson (1999) for additional details on these models.

8.1 Simple Markov switching

Here the switching is under the control of a Markov-chain updating mechanism with fixed transition probabilities. Let $f(Y_t|S_t = j, \Psi_{t-1})$ denote the probability density of the dependent variable at time t when regime j is operating, where Ψ_{t-1} represents the history of the process to date t–1, and let the probability of falling in regime j at time t evolve according to

$$\Pr(S_t = j \mid \Psi_t) = \frac{f(Y_t \mid S_t = j, \Psi_{t-1}) \Pr(S_t = j \mid \Psi_{t-1})}{\sum_{i=1}^{M} f(Y_t \mid S_t = i, \Psi_{t-1}) \Pr(S_t = i \mid \Psi_{t-1})} \tag{8.2}$$

where

$$\Pr(S_t = j \mid \Psi_{t-1}) = \sum_{i=1}^{M} p_{ji} \Pr(S_{t-1} = i \mid \Psi_{t-1}) \tag{8.3}$$

The transition probabilities $p_{ji} = \Pr(S_t = j \mid S_{t-1} = i)$ are $M(M-1)$ fixed parameters to be estimated, subject to $\sum_{j=1}^{M} p_{ji} = 1$. The likelihood function maximized for estimation is

$$L_T = \sum_{t=1}^{T} \log \sum_{j=1}^{M} f(Y_t \mid S_t = j, \Psi_{t-1}) \Pr(S_t = j \mid \Psi_{t-1}) . \tag{8.4}$$

Note that the transition probabilities are mapped onto the real line for unrestricted estimation, so that the parameters are t_{ji} such that

$$p_{ji} = \frac{\exp(t_{ji})}{1 + \sum_{k=1}^{M-1} \exp(t_{ki})}, j = 1,\ldots,M-1 \text{ and } p_{Mi} = 1 - \sum_{j=1}^{M-1} p_{ji}, i = 1,\ldots,M. \tag{8.5}$$

The t_{ji} are also reported in the output, with standard errors. Standard errors are not reported for the p_{ji}.

The series $\Pr(S_t = j|\Psi_t)$ for $j = 1,\ldots, M-1$, the 'filter probabilities', are a by-product of the estimation. From these we can also compute the sequences of 'smoothed probabilities' $\Pr(S_t = j \mid \Psi_T)$, which are obtained from the backwards recursion

©James Davidson 2015

$$\Pr(S_t = j \mid \Psi_T) = \sum_{i=1}^{M} p_{ij} \frac{\Pr(S_t = j \mid \Psi_t)\Pr(S_{t+1} = i \mid \Psi_T)}{\Pr(S_{t+1} = i \mid \Psi_t)}, \quad t = T-1,\ldots,1. \tag{8.6}$$

These series should give the best indication of which regime the system is occupying at each date in the sample; see Kim and Nelson (1999) Section 4.3.1 for details.

8.2 Explained Switching

This setup is the same as Case 1 except that the t_{ji} are functions of specified predetermined variables. Two versions of this model can be specified. In the first, the equations have the general form

$$t_{ji} = \alpha_j + \beta_{ji} D_i + \sum_{m=1}^{R} \gamma_{jm} x_{m,t-J}, \quad j = 1,\ldots,M-1 \tag{8.7}$$

where $D_i = 1$ if the current regime is i, for $i = 2,\ldots,M$ and zero otherwise. $J \geq 0$ is a fixed lag that can be specified interactively, although distributed lags would need to be set up manually. In other words, the representation may depend on the current regime through possible intercept shifts. This model nests the Markov switching model as the case where $\gamma_{jm} = 0$ for all j and m. However, the way the probability of switching to regime j depends on x_{mt} does not depend on i.

In the second, extended version of the model, called "regime-dependent coefficients", the equations take the form

$$t_{ji} = \alpha_{ji} + \sum_{m=1}^{R} \gamma_{jim} x_{m,t-J}, \quad j = 1,\ldots,M-1,\ i = 1,\ldots,M \tag{8.8}$$

Note that equation (8.8) nests the model in (8.7), which can be obtained by imposing suitable restrictions. However, the simpler model is specified slightly differently, and so is retained for compatibility. In (8.7) the dummy-variable format estimates the intercepts for regimes $i = 2,\ldots,M$ in the form $\alpha_j + \beta_{ji}$, whereas in (8.8) the equivalent parameter is α_{ji}. However, note that optionally any switching parameter for regimes $i = 2,\ldots,M$ can be expressed as differences from regime 1, and hence the β_{ji} can also be estimated in this setup.

In the current version of the program, the vector x_{mt} may include the dependent variable(s), although in this case the condition $J \geq 1$ is enforced. Model (8.7) nests cases such as the threshold autoregression (TAR) model. Put $M = 2$ and $R = 1$, suppress the dummies, and (for example) set $x_{1t} = y_{t-1} - c$ (or alternatively $x_{1t} = |y_{t-1}| - c$) for a threshold parameter c. For large γ_{11} this two-regime model is arbitrarily close to the case where $p_{1i} = I(y_{t-1} > c)$ and $p_{2i} = I(y_{t-1} \leq c)$ where $I(.)$ denotes the indicator function. The parameters could be imposed, or alternatively estimated as $c = -\alpha_1/\gamma_{11}$.

8.3 Hamilton's Markov switching model

Here, the mean and/or unconditional variance in a finite-order AR-ARCH or VAR-VARCH process depends on the regime under which the variables are generated. The model has the representation

$$Y_t - \mu(S_t) = \phi_1(Y_{t-1} - \mu(S_{t-1})) + \ldots + \phi_p(Y_{t-p} - \mu(S_{t-p})) + u_t \tag{8.9}$$

Where $u_t = \sigma(S_t)h_t^{1/2}e_t$ and

$$h_t = 1 + \alpha_1 u_{t-1}^2 / \sigma^2(S_{t-1}) + \cdots + \alpha_p u_{t-p}^2 / \sigma^2(S_{t-p}). \tag{8.10}$$

©James Davidson 2015

Here, $\mu(1),\ldots,\mu(M)$ and $\sigma^2(1),\ldots,\sigma^2(M)$, as well as transition probabilities p_{ji}, are parameters of the process. The mechanism governing the updating of probabilities generalizes (8.2)-(8.4) to incorporate M^{p+1} distinct states, in which each of the p lag terms fall in each of M regimes. See Hamilton (1989), and Hamilton and Susmel (1994) for details.

Note that their row sums are constrained to lie in [0,1], but for estimation purposes they are mapped into the real line for unconstrained estimation. In other words, the fitted parameters are r_{ji} where

$$p_{ji} = \frac{\exp(r_{ji})}{1+\sum_{m=1}^{M-1}\exp(r_{mi})} \quad j = 1,\ldots,M-1 \text{ and } \quad p_{Mi} = \frac{1}{1+\sum_{k=1}^{M-1}\exp(r_{ki})}$$

8.4 The Smooth Transition (ST) Model

This is an alternative regime-switching specification, (see Granger and Teräsvirta 1994, Teräsvirta 1998) comparable to explained Markov switching, but without an explicitly probabilistic interpretation.

8.4.1 Single Transition Model

Let equation (4.1) be written in the implicit form

$$u_t = u(Y_{t-s}, \boldsymbol{x}_{t-s}, s \geq 0; \boldsymbol{\psi}) \tag{8.11}$$

where $\boldsymbol{x}_t = (\boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}, \boldsymbol{x}_{3t}, \boldsymbol{x}_{7t})$ and $\boldsymbol{\psi}$ here denotes the complete set of model parameters. If $\boldsymbol{\psi}_1$ and ψ_2 represent alternative parameter values and $u_{jt} = u(\boldsymbol{\psi}_j)$ for $j = 1,2$, the ST model takes the form

$$u_t = G_t\, u_{1t} + (1 - G_t)\, u_{2t} \tag{8.12}$$

where

$$G_t = \frac{1}{1+\exp\{\gamma(z_t - \tau - \boldsymbol{\pi}_8'\boldsymbol{x}_{8t})\}}. \tag{8.13}$$

so that $0 \leq G_t \leq 1$. This specification includes the well-known SETAR model as a special case, by allowing autoregressive parameters to switch, with $z_t = Y_{t-d}$. However, as in the Markov-switching models, any desired subset of the parameters ψ can be allowed to vary across regimes driven by any variable(s). An ST-GARCH model can also be specified, by replacing $h_t = h(Y_{t-s}, \boldsymbol{x}_{t-s}, s \geq 1; \psi)$ in (6.1) or (6.2) by

$$h_t = G_t h_{1t} + (1 - G_t)\, h_{2t} \tag{8.14}$$

where $h_{jt} = h(\psi_j)$ for $j = 1,2$.

With $\boldsymbol{\pi}_8 = 0$, the intercept τ measures the transition value of the regime 'indicator' z_t, whose location determines which regime receives most weight at date t. With $\gamma > 0$, for example, regime 1 dominates when $z_t < \tau$ and regime 2 dominates when $z_t > \tau$. However, the switch value can depend on additional explanatory variables $\boldsymbol{x}_{8t}$, if desired. As the 'smoothness' parameter γ takes (absolutely) large values the transition becomes abrupt, and the model is similar to the ordinary threshold (TAR-type) models. *Note*: changing the sign of γ replaces G_t with $(1 - G_t)$, and so is equivalent to interchanging the values of $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$. The model has two observationally equivalent versions. Be careful to interpret the estimates correctly.

©James Davidson 2015

8.4.2 Double Transition Model

A further variant is the 'double transition' model, in which

$$G_t = \frac{1}{1+\exp\{\gamma(z_t - \tau_1 - \boldsymbol{\pi}_8'\boldsymbol{x}_{8t})(z_t - \tau_2 - \boldsymbol{\pi}_9'\boldsymbol{x}_{9t})\}}. \tag{8.15}$$

With $\boldsymbol{\pi}_8 = \mathbf{0}$, $\boldsymbol{\pi}_9 = \mathbf{0}$ and assuming $\tau_1 < \tau_2$ (say), this model can be seen as defining an interval $[\tau_1 , \tau_2]$. With $\gamma > 0$, the model is predominantly in regime 1 when $\tau_1 \le z_t \le \tau_2$ (such that the multiplied factors have different signs) and in regime 2 otherwise.

8.4.3 Structural Change (Date-Transition) Model

An important application of ST models is to capture structural breaks, and estimate break-dates. Set $z_t = t$ (the time trend), and then with $\pi_7 = 0$, τ measures the break-date between consecutive regimes. γ can be fixed at a large value to create a sudden break, or can permit smooth variation.

8.5 Testing for Breaks and Regimes

One problem with Markov and ST models is that the switch probabilities, or switch parameters, are unidentified when there are in fact no differences between the models in each regime. This fact can give rise to numerical problems in optimizing the models. It also means that it is difficult to test the hypothesis of "no switching", since parameters are unidentified under the null hypothesis. The latter problem may be resolvable by computing "sup-*F*" tests, based on the largest value of the statistic over the range of eligible values of the unidentified parameter. These statistics can be computed using TSM's Criterion Plot feature.

For example, the sup-Wald test of Andrews (1993) can be computed in this way. Use the Date-Transition model to parameterize the switches, select the 'Estimate Regime Differences' option, and set up a Wald test of the significance of some/all the Regime 2 parameters. The sup-Wald statistic is computed by fixing γ at a large value (say 100) and generating a grid of values of the switch-date parameter τ. The maximum of the statistic over the grid is reported in the output.

Also see Sections 12.6 and 12.7 on pre-programmed tests of specification using the 'sup' principle.

©James Davidson 2015

9 Post-Estimation Options

9.1 Residuals and Associated Series

Two varieties of residual can be plotted and retrieved for further analysis. In the case of the models represented in (4.1), (4.15)+(4.16) and (5.1), these are respectively the ordinary residuals $\hat{u}_t$ and the (variance-)adjusted residuals $\hat{h}_t^{-1/2}\hat{u}_t$, where the hats denote evaluation at the estimated parameters. If there is no conditional variance model specified, adjustment simply means that the variance is normalized to unity. Otherwise the conditional variances $\hat{h}_t$ are also reported. In systems, the adjusted residuals are $\hat{\boldsymbol{H}}_t^{-1/2}\hat{\boldsymbol{u}}_t$.

In probit and logit models, the ordinary residuals are computed as $\hat{u}_t = Y_t - F(\hat{h}_t^{-1/2}\hat{z}_t)$, while in the probit models the 'adjusted' residuals are what are normally called the generalized residuals, defined as

$$\hat{e}_t = \frac{\phi(\hat{h}_t^{-1/2}\hat{z}_t)(Y_t - F(\hat{h}_t^{-1/2}\hat{z}_t))}{F(\hat{h}_t^{-1/2}\hat{z}_t)(1-F(\hat{h}_t^{-1/2}\hat{z}_t))} \tag{9.1}$$

where ϕ denotes the standard normal density. This is the series which is orthogonal by construction to the regressors, or more generally to the derivatives of $\hat{h}_t^{-1/2}\hat{z}_t$, evaluated at the maximum likelihood estimates.

In the ordered probit and logit models, the ordinary residuals are computed as

$$\hat{u}_t = Y_t - E(Y_t \mid x_{1t},\ldots,x_{7t}) = Y_t - \sum\nolimits_{j=1}^{J-1} j\hat{F}_j - J(1-\hat{F}_{J-1}). \tag{9.2}$$

where $\hat{F}_j = F(\hat{\gamma}_1 + \cdots + \hat{\gamma}_j - \hat{h}_t^{-1/2}\hat{z}_t)$. whereas the generalized residuals are again the series that is orthogonal by construction to the derivatives of $\hat{z}_t$ or $\hat{h}_t^{-1/2}\hat{z}_t$. These are defined as

$$\hat{e}_t = 1_{\{Y_t=0\}}\frac{\hat{\phi}_0}{\hat{F}_0} + \sum_{j=1}^{J-1} 1_{\{Y_t=j\}}\frac{\hat{\phi}_j - \hat{\phi}_{j-1}}{\hat{F}_j - \hat{F}_{j-1}} - 1_{\{Y_t=J\}}\frac{\hat{\phi}_{J-1}}{1-\hat{F}_{J-1}} \tag{9.3}$$

where $1_{\{Y_t=j\}}$ is the indicator function, equal to 1 when the argument is true and 0 otherwise, $\hat{F}_0 = F(-\hat{h}_t^{-1/2}\hat{z}_t)$, for $j = 1,\ldots,J-1$, and the $\hat{\phi}_j$ are the corresponding density functions evaluated at the same points. Note that (9.3) reduces to (9.1) in the binary data case.

In count data models the residuals are constructed as $\hat{u}_t = Y_t - \hat{\phi}_t$.

In all these cases, the adjusted/generalized residuals are used to compute the various diagnostic tests available, using the LM and conditional moment principles.

In switching models, the residuals are computed for each regime, and the reported series are the weighted average of these, where the weights are the filter probabilities. In switching models the series of conditional probabilities are also retrievable, including the smoothed probabilities computed using Kim's algorithm (see Kim and Nelson 1999) and the variable switch probabilities in the case of explained switching models; specifically if $p_{ji}(t)$ denote the time-varying transition probabilities, the series available for plotting are of the form $\overline{p}_j(t) = \sum_{i=1}^{M} p_{ji}(t)$.

©James Davidson 2015

9.2 Model Performance and Selection Criteria

The program output routinely reports the following statistics.

1. By default the maximum value of the maximand is reported, that is, $-\hat{C}_T$ where estimation is by minimizing the criterion C_T. This criterion may be a sum of squared residuals, or quadratic form in methods of moments estimation. Of course, in maximum likelihood estimation $-\hat{C}_T = \hat{L}_T$, where $\hat{L}_T$ is the maximum of the log-likelihood function defined by the relevant formula; see equations (10.3)–(10.21) below.

   Optionally, the sign can be reversed, so that the minimand is reported. It appears most natural to report the maximand when estimation is by maximum likelihood, but possibly less confusing to report the minimand in method of moments estimation, when there is no natural analogue of the log-likelihood function to consider.

2. By default, the following model selection criteria are reported where in each case *p* is the number of fitted parameters, and *T* is sample size.

   Akaike Information Criterion (AIC): $-C_T - p$

   Schwarz information criterion (SIC): $-C_T - \frac{1}{2} p \log T$

   Hannan-Quinn criterion (HQC): $-C_T - p \log\log T$

   Note that by default, each of these measures provides a basis for choosing between a set of alternative models on the criterion "Larger = Better".

   By choosing to report the minimand, the signs of all these criteria can be optionally reversed, so that the selection criterion should be "Smaller = Better". Note that econometrics packages often report the criteria in "Smaller = Better" style and also divide them conventionally by *T*.

   The SIC and the HQC are consistent selection criteria, meaning that if the true model is one of those compared, it will be selected with a probability that approaches 1 as *T* increases. This is not true of the AIC.

3. Sum of squared residuals ("Sum of Squares", or RSS).

4. R^2 ("R-squared") is defined, in every case, as the square of the correlation coefficient of the actual and fitted values of the dependent variable. Note that it is, accordingly, always defined on the interval [0, 1]. This formula agrees with the conventional formula, 1 – RSS/TSS where TSS is the sum of squared mean deviations of the dependent variable, in the case of linear least squares estimation with an intercept included. However, note that the latter formula does not generalize to nonlinear models, IV estimates, etc. The given formula is always valid, although not, of course, a consistent model selection criterion.

5. $\bar{R}^2$ ("R-bar-squared") = $1-(1-R^2)(T-1)/(T-p))$ where *p* is the number of parameters fitted in the equation. This is a commonly used model selection criterion for regression models.

6. The standard deviation (SD), skewness (Sk) and kurtosis (Kt) of the residuals, and Jarque-Bera statistic. The latter is defined as

$$\mathrm{JB} = \frac{T}{6}\left(\frac{\mathrm{Sk}}{\mathrm{SD}^3}\right)^2 + \frac{T}{24}\left(\frac{\mathrm{Kt}-3\mathrm{SD}^4}{\mathrm{SD}^4}\right)^2 \tag{9.4}$$

©James Davidson 2015

and is asymptotically distributed as chi-squared with two degrees of freedom when the residuals are normal (Gaussian) and independently distributed.

In addition to these standard statistics, some special statistics are reported for particular models.

7. For discrete data models the Likelihood Ratio index is reported, defined as

$$\text{LRI} = 1 - \hat{L}_T / L_{0T} \tag{9.5}$$

where L_{0T} is the log-likelihood function evaluated with all parameters except for the equation intercept (and in the Markov-switching case, the switching probabilities) replaced by 0. The LRI statistic tends to 0 when the explanatory variables have true zero coefficients, and it cannot exceed 1, noting that both functions are negative, being sums of log-probabilities, and that $\hat{L}_T \geq L_{0T}$.

8. For instrumental variables and GMM estimation, the Sargan overidentification test and the Durbin-Wu-Hausman exogeneity test are reported. See (e.g.) Davidson (2000) Chapter 8 for details of these tests.

Statistics commonly reported by other regression packages, but available only as options in TSM, are the Durbin-Watson statistic and the *F* statistic for joint significance of the slope coefficients, sometimes called the "F-test of the Regression". The version of the latter test implemented in TSM excludes trend and seasonal dummies, and also lagged dependent variables, from the tested set. Therefore, it tests only the joint significance of true exogenous variables in the regression.

9.3 Q Tests

Standard tests for serial dependence are the provided by the *Q* tests for levels (Box-Pierce 1970) and squares (McLeod-Li 1983) of the data. The latter is a test for nonlinear dependence in a serially uncorrelated (white noise) series. The default statistic is the Box-Pierce (1970) formula

$$Q(m) = n\sum_{j=1}^{m} r_j^2 \tag{9.6}$$

which is asymptotically chi-squared with $m - p - q$ degrees of freedom when applied to the residuals of an ARMA(p, q) model. Optionally, this can be replaced by asymptotically equivalent Ljung-Box (1978) formula

$$Q(m) = n(n+2)\sum_{j=1}^{m} \frac{r_j^2}{n-j} \tag{9.7}$$

with claimed better small-sample properties.

9.4 Ex-post forecasts

One-step ex-post forecasts are obtained by fitting the model using data up to time T, and then computing the usual fitted equation and residuals for periods $T + 1$ to $T + F$, such that all right-hand side variables are treated as known. The main purpose of this option is to test model stability.

In the general case, two test statistics are computed,

©James Davidson 2015

$$\text{Forecast Test I} = \frac{\sum_{t=T+1}^{T+F} \hat{u}_t^2}{Var_T(\hat{u}_t)} \tag{9.8}$$

where the denominator is the usual residual variance from the sample period, and

$$\text{Forecast Test II} = \frac{F^{-1}\sum_{t=T+1}^{T+F} \hat{u}_t^2 - T^{-1}\sum_{t=1}^{T} \hat{u}_t^2}{\sqrt{F^{-1}\text{Var}_F(\hat{u}_t^2) + T^{-1}\text{Var}_T(\hat{u}_t^2)}} \tag{9.9}$$

Test I is an asymptotically valid version of Chow's prediction test, distributed as $\chi^2(F)$ under the null hypothesis of model stability, *also* assuming the disturbances are Gaussian. Test II is the usual difference-of-means test on the residual and forecast error variances. It is asymptotically N(0,1) under the stability hypothesis, assuming 4th moments exist, where 'asymptotic' is interpreted as $\min(F,T) \to \infty$. That is, these tests are appropriate to 'small' and 'large' forecast periods, respectively.

In single equation linear regression models, Forecast Test II is replaced by Chow's stability test, with the formula (*F* version)

$$\text{Chow} = \frac{T+F-2k}{k} \frac{\sum_{t=1}^{T+F} \dot{u}_t^2 - \sum_{t=T+1}^{T+F} \tilde{u}_t^2 - T^{-1}\sum_{t=1}^{T} \hat{u}_t^2}{\sum_{t=T+1}^{T+F} \tilde{u}_t^2 + T^{-1}\sum_{t=1}^{T} \hat{u}_t^2} \tag{9.10}$$

where $\tilde{u}_t$ and $\dot{u}_t$ represent respectively residuals from the model fitted to the forecast period and the whole period, and *k* is the number of regression parameters. The Chi-squared version of the test omits the *k* in the denominator. The theory of this test is well known. (If it is desired to compute formula (9.9) instead, compute the regression through the dynamic equation dialog).

9.5 Ex-ante Multi-step Forecasts

Multi-step ex-ante forecasts can be computed beyond the end of the available data period provided the model does not contain exogenous variables (other than the trend dummy or GARCH-M term). Two methods are available.

9.5.1 Analytic Forecasts

These are computed by solving the dynamic model forward with zero shocks, to generate the expected path. Confidence bands are computed based on the standard error formula

$$S_{T+k} = \sqrt{\sum_{j=0}^{k} \varphi_j^2 s_{T+k-j}^2} \tag{9.11}$$

where $\varphi(L) = (1-L)^{-d_1}\,\phi(L)^{-1}\theta(L)$, and s_t^2 is the predicted conditional variance of the disturbance at date *t*. These are asymptotically valid assuming independent innovations, but ignore parameter uncertainty, s_t^2 is estimated by the equation residual variance in models without conditional heteroscedasticity, and otherwise is the multi-step variance forecast generated from equations (6.1) or (6.2), as appropriate. Note that in case (6.2), the unbiased forecast of $\log h_t$ is generated, and hence the implied forecast of h_t is biased towards zero (see Nelson 1991 for details). The reported bands should be treated as lower bounds on the 2-standard error bands in this case.

Forecasts are computed by default using a numerical solution to the model simulation algorithm using zero shocks (in contrast to stochastic simulation used for Monte Carlo

©James Davidson 2015

forecasts) and hence are available for both linear and nonlinear structures. The exception is Markov-switching models (see below).

9.5.2 Moving Average Coefficients (Impulse and Step Responses)

There is an option to compute the sequence of solved moving average coefficients (impulse responses) and also the cumulated sequence (step responses). In the case of equation (4.1), for example, the sequence is the solved coefficients of the polynomial

$$\phi(L)^{-1}\theta(L)(1-L)^{-d_1}. \tag{9.12}$$

In the case of equations (6.1) or (6.2), the sequence is the solved coefficients of

$$1-\beta(L)^{-1}\delta(L)(1+\alpha((1-L)^{d_2}-1)) \tag{9.13}$$

Since the weights are computed by perturbing the numerical solution of the model simulation algorithm, results are also available for nonlinear dynamic structures such as bilinear, error-correction and smooth transition models. In the case of a nonlinear-in-variables specification, the sequence of weights so produced represents a linear approximation to the actual expected response to a unit shock. This option is not available for Markov-switching models (see below).

9.5.3 Forecast Error Variance Decomposition

For forecasts of multi-equation models using the analytic approach, error variance decompositions using the method of Lütkepohl (2007) Section 2.3.3 are optionally available. This output shows, for each forecast horizon, the proportion of the forecast error variance attributable to the shocks to each equation in the system. This option is not available for Markov-switching models.

9.5.4 Forecasting Regime-Switching Models

The following formulae are used for Markov-switching models. Let $\hat{y}_{T+k}(j_1,...,j_M)$ denote the optimal forecast of y_{T+k} for the case where it is known that the process is in regime j_i in period $T+i$, $i = 1,\ldots,K$. Then the K-step forecast $\hat{y}_{j,T+K}$ is computed as

$$\hat{y}_{T+K}=\sum_{j_1=1}^{M}\cdots\sum_{j_K=1}^{M}\hat{P}_{T+1,j_1}\cdots\hat{P}_{T+K,j_K}\hat{y}_{T+K}(j_1,...,j_M) \tag{9.14}$$

where

$$\hat{P}_{T+t,j}=\Pr(S_{T+t}=j\mid\Psi_T)=\sum_{i=1}^{M}p_{ji}\hat{P}_{T+t-1,i}\,,\quad K=1,2,\ldots,F. \tag{9.15}$$

Note that (9.14) can be computed by a straightforward recursion, and does not in practice involve M^K terms. The sequence in (9.15) converges to the ergodic (steady state) probabilities if the p_{ji} are constants, while under explained switching (case 2) the forcing variables are treated as exogenously given for the purposes of the forecast. In Hamilton's model (case 3) the formula is modified as appropriate to allow for the M^{p+1} distinct states. Confidence bands can be computed using the standard error formula

$$S_{T+K}=\sqrt{\sum_{j_1=1}^{M}\cdots\sum_{j_k=1}^{M}\hat{P}_{T+K,j_1}\cdots\hat{P}_{T+K,j_k}(S^2_{T+K}(j_1\cdots j_k)+\hat{y}^2_{j,T+K}(j_1\cdots j_k))-\hat{y}^2_{T+K}}\,,. \tag{9.16}$$

$K = 1, 2, \ldots, F$, where $S_{T+K}(j_1,\ldots,j_k)$ denotes the K-step forecast standard error analogous to (9.11), conditional on regime configuration $j_1,\ldots,j_k$. This formula takes account of the uncertainty about which regime an observation represents, but ignores parameter uncertainty. There is no short cut to evaluation of all the M^K terms, hence it can be

©James Davidson 2015

computationally intensive for large *M* and/or *K*, although if it converges rapidly, it can be extrapolated. See Davidson (2004b) for additional details.

This method is currently enabled only for ARMA/ARIMA/ARFIMA models with ARCH/GARCH conditional heteroscedasticity. For cases of bilinear and ECM structures with Markov switching, use the Monte Carlo forecasting option.

9.5.5 Monte Carlo Forecasts

Monte Carlo forecasts are available for all models, linear and nonlinear. Here the dynamic model is stochastically simulated *F* steps forward using one of three options for generating the shocks: Gaussian, likelihood matching, and bootstrap. This method is available for all the models. For some models, such as discrete and count data, smooth transition models and user-supplied models, Monte Carlo is the only forecasting option implemented.

There are two options for reporting point forecasts; as either the means of the Monte Carlo distributions for each step ahead, or as the medians of these distributions. In the former case, 2-standard error bands are constructed using the variances of the Monte Carlo distributions. In the latter case, the 2.5% and 97.5% quantiles are reported, to provide an approximate 95% confidence band. Note that the validity of the median forecasts and confidence bands does not depend on the forecast mean and variance being well-defined.

©James Davidson 2015

10 Estimation Criteria

10.1 Single Equation Methods

10.1.1 Least Squares

For equations (4.1)+(4.2) only, $C_T = \sum_{t=1}^{T} u_t^2$.

10.1.2 Instrumental Variables

For equations (4.1)+(4.2), the initial estimation minimand is

$$C_T = \left(\sum_{t=1}^{T} \boldsymbol{z}_t' u_t \right) \left(\sum_{t=1}^{T} \boldsymbol{z}_t \boldsymbol{z}_t' \right)^{-1} \left(\sum_{t=1}^{T} \boldsymbol{z}_t u_t \right) \tag{10.1}$$

where $\mathbf{z}_t$ is a vector of instruments selected from the data set. If the covariance matrix formula is set to Robust (the default) or HAC (see below) then GMM can be computed as a two- or multi-stage estimator. The first run minimises (10.1). Further runs then minimise

$$C_T = \left(\sum_{t=1}^{T} \boldsymbol{z}_t' u_t \right) \hat{\boldsymbol{M}}^{-1} \left(\sum_{t=1}^{T} \boldsymbol{z}_t u_t \right) \tag{10.2}$$

where $\hat{\boldsymbol{M}} / T$ is the estimated covariance matrix of $\mathbf{z}_t u_t$, either White or HAC, according to the current selection, with the residuals evaluated at the last-stage estimates. The kernel and bandwidth choice for the HAC estimator is the same as currently selected for the computation of tests and standard errors. (Note: multi-stage estimation is disabled under criterion grid plotting and multiple ARMA estimation.)

10.1.3 Gaussian maximum likelihood

e_t is standard Gaussian. The criterion function maximised is

$$L_T = -\frac{T}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \left(\log h_t + \frac{u_t^2}{h_t} \right) \tag{10.3}$$

10.1.4 Student-*t* maximum likelihood

e_t is Student's *t*-distributed with $\nu > 2$ degrees of freedom. The criterion is

$$L_T = T \log \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} - \frac{T}{2} \log \pi(\nu-2) - \frac{1}{2} \sum_{t=1}^{T} \left(\log h_t + (\nu+1) \log \left(1 + \frac{u_t^2}{(\nu-2) h_t} \right) \right) \tag{10.4}$$

To improve numerical stability, the parameter actually estimated by TSM is $\nu^{1/2}$.

10.1.5 Skew-student maximum likelihood

e_t has the skewed Student's *t* distribution, with parameters $\nu > 2$ and $\xi > 0$. (See Fernandez and Steel (1998), Lambert and Laurent (2001)). The criterion is

$$\begin{aligned} L_T = {} & T \log \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} - \frac{T}{2} \log \pi(\nu-2) + T \log \left(\frac{2s}{\xi + \xi^{-1}} \right) \\ & - \frac{1}{2} \sum_{t=1}^{T} \left(\log h_t + (\nu+1) \log \left(1 + \frac{(s h_t^{-1/2} u_t + m)^2}{\nu - 2} \xi^{-2 I_t} \right) \right) \end{aligned} \tag{10.5}$$

©James Davidson 2015

where

$$m=\frac{\Gamma((\nu-1)/2)(\nu-2)^{1/2}}{\Gamma(\nu/2)\sqrt{\pi}}\left(\xi-\frac{1}{\xi}\right),\ s=\sqrt{\xi^2+\xi^{-2}-1-m^2}$$

and

$$I_t=\begin{cases}1, & sh_t^{-1/2}u_t+m\geq 0\\ -1, & \text{otherwise.}\end{cases}$$

The parameter ξ measures the skewness of the distribution, which reduces to (10.4) when $\xi = 1$.

10.1.6 GED maximum likelihood

e_t has the Generalized Error Distribution (GED) with parameter $\nu > 0$ (see Nelson 1991). The criterion is

$$L_T=T\log\frac{\nu 2^{(1+1/\nu)}}{\lambda_\nu\Gamma(1/\nu)}-\frac{1}{2}\sum_{t=1}^{T}\left(\log h_t+\left|\frac{u_t}{\lambda_\nu h_t^{1/2}}\right|^\nu\right) \tag{10.6}$$

where

$$\lambda_\nu=\sqrt{\frac{\Gamma(1/\nu)2^{-2/\nu}}{\Gamma(3/\nu)}}\ .$$

Note that GED corresponds to the Gaussian case when $\nu = 2$, and is leptokurtic when $\nu < 2$.

10.1.7 Whittle maximum likelihood

For equation (4.3) with Gaussian errors only. The criterion function (see Hauser 1999) is

$$L_T=M\left(\log 2\pi-1-\log\frac{1}{M}\sum_{j=1}^{M}\frac{I_j}{g_j}\right)-\sum_{j=1}^{M}\log g_j \tag{10.7}$$

where I_t is the jth point of the periodogram of Y_t (expressed in mean deviations), $M = [T/2]$, and

$$g_j=\left(2-2\frac{\cos 2\pi j}{T}\right)^{-d}\frac{\left(\sum_{m=0}^{q-1}\theta_{m+1}\cos 2\pi mj/T\right)^2+\left(\sum_{m=0}^{q-1}\theta_{m+1}\sin 2\pi mj/T\right)^2}{\left(\sum_{m=0}^{p-1}\phi_{m+1}\cos 2\pi mj/T\right)^2+\left(\sum_{m=0}^{p-1}\phi_{m+1}\sin 2\pi mj/T\right)^2} \tag{10.8}$$

This algorithm is quite a lot faster than conditional least squares (i.e. formula (10.3) without GARCH) in large samples.

10.1.8 Probit and Logit

The log-likelihood function is defined as

$$L_T=\sum_{t=1}^{T}\left[Y_t\log F(Y_t^*)+(1-Y_t)\log(1-F(Y_t^*))\right] \tag{10.9}$$

where Y_t denotes the binary (0 or 1) dependent variable, $F()$ denotes the standard normal or logistic CDF, and Y_t^* is defined in (4.33) and (6.8) or (6.9).

©James Davidson 2015

10.1.9 Ordered Probit and Logit

Letting $1_{\{Y_t=j\}}$ denote the indicator function, taking the value 1 when its argument is true and 0 otherwise, the log-likelihood for the model of J states ($J \geq 3$) can be written as

$$\begin{aligned} L_T = \sum_{t=1}^{T} \Bigg[& 1_{(Y_t=0)} \log F(-Y_t^*) + \\ & \sum_{j=1}^{J-1} 1_{(Y_t=j)} \log \left\{ F\left(\sum_{i=0}^{j} \gamma_j - Y_t^*\right) - F\left(\sum_{i=0}^{j-1} \gamma_j - Y_t^*\right) \right\} + \\ & 1_{(Y_t=J)} \log \left\{ 1 - F\left(\sum_{i=0}^{J-1} \gamma_j - Y_t^*\right) \right\} \Bigg] \end{aligned} \tag{10.10}$$

where $\gamma_0 = 0$ and F and Y_t^* are defined as before. Note that for $1 \leq j \leq J-1$, $\gamma_j = 0$ is a feasible value only if the jth category is empty. Then, the corresponding term is omitted from the likelihood function.

10.1.10 Poisson

For count data, the log-likelihood in the Poisson model is defined as

$$L_T = \sum_{t=1}^{T} \left[-\phi_t + Y_t \log \phi_t - \log \Gamma(1 + Y_t) \right] \tag{10.11}$$

where ϕ_t is defined in (4.35) and $\Gamma(.)$ represents the gamma function. In the negative binomial models, the log-likelihood is

$$L_T = \sum_{t=1}^{T} \left[\log\left(\frac{\Gamma(Y_t + \nu_t)}{\Gamma(Y_t + 1)\Gamma(\nu_t)} \right) + \nu_t \log\left(\frac{\nu_t}{\nu_t + \phi_t} \right) + Y_t \log\left(\frac{\phi_t}{\nu_t + \phi_t} \right) \right] \tag{10.12}$$

where $\nu_t = 1/\alpha$, and $\nu_t = \phi_t / \alpha$, in the first and second cases, respectively. Note that in either case, (10.12) reduces to (10.11) when $\alpha = 0$.

10.1.11 Zero-inflated Discrete

Let the expressions in (10.10), (10.11) and (10.12) be denoted generically as

$$L_T = \sum_{t=1}^{T} l_{1t} \,. \tag{10.13}$$

The zero-inflated models introduce a distribution function $F_2(w_t)$ representing the probability of the 'zero' regime, and define the log-likelihood as

$$L_T = \sum_{t=1}^{T} \log \left\{ F_2(w_t) 1_{(Y_t=0)} + (1 - F_2(w_t)) \exp(l_{1t}) \right\} \tag{10.14}$$

where $1_{(.)}$ denotes the indicator function of its argument.

10.2 System Methods

10.2.1 Least Generalized Variance

For a system of equations without simultaneity, conditional variance model or regime switching, the basic minimand is the generalized variance (determinant of the covariance matrix). The criterion maximized is therefore

©James Davidson 2015

$$L_T = -\frac{T}{2}\log\det\left(T^{-1}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{u}_t'\right). \tag{10.15}$$

Least Generalized Variance (LGV) is equivalent to conditional Gaussian ML for these models.

For a simultaneous equations model in which dependent variables are also specified as explanatory variables, such that $\boldsymbol{B} \neq \boldsymbol{I}$ in (5.1), the maximand is

$$L_T = T\log|\det\boldsymbol{B}| - \frac{T}{2}\log\det\left(T^{-1}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{u}_t'\right). \tag{10.16}$$

10.2.2 Generalized Method of Moments

To estimate a system of equations by GMM, the minimand can be written by expressing the model in stacked form. Let $\boldsymbol{u}_j = (u_{j1},\ldots,u_{jT})'$ for $j = 1,\ldots,N$ and $\boldsymbol{u}^* = \text{vec}(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_N)$. Also let $\boldsymbol{W} = \boldsymbol{I}_N \otimes \boldsymbol{Z}$ where $\boldsymbol{Z}$ is the matrix of instruments specified by the user. The basic GMM minimand is then

$$C_T = \tfrac{1}{2}\boldsymbol{u}^{*\prime}\Lambda^{-1}\boldsymbol{W}(\boldsymbol{W}'\Lambda^{-1}\boldsymbol{W})^{-1}\boldsymbol{W}'\Lambda^{-1}\boldsymbol{u}^* \tag{10.17}$$

where $\Lambda = \boldsymbol{S} \otimes \boldsymbol{I}_T$. $\boldsymbol{S}$ is the estimator of the error variance matrix, which on the first run of this estimator is replaced by the $N\times N$ identity matrix. However, if the estimation is iterated (giving the "Run" command again immediately, without any intervening user actions) the optimization is repeated with $\boldsymbol{S}$ containing the estimate of the covariance matrix from the current residuals. This iteration can be repeated as often as desired.

Further, if the covariance matrix option is set to Robust (the default) or HAC (see below) then the efficient GMM minimand takes the form

$$C_T = \tfrac{1}{2}\boldsymbol{u}^{*\prime}\Lambda^{-1}\boldsymbol{W}\hat{\boldsymbol{M}}^{-1}\boldsymbol{W}'\Lambda^{-1}\boldsymbol{u}^* \tag{10.18}$$

where $T^{-1}\hat{\boldsymbol{M}}$ denotes, respectively, either the White-Eicker or the Newey-West estimator of the covariance matrix of the vectors $z_k^* u_k^*$ for $k = 1,\ldots,NT$ where $z_k^{*\prime}$ denotes the row of the matrix $\Lambda^{-1}\boldsymbol{W}$. The kernel and bandwidth choice for the HAC estimator is the same as currently selected for the computation of tests and standard errors.

10.2.3 Gaussian ML with Conditional Heteroscedasticity

To estimate a Gaussian system featuring a conditional variance model, the maximand takes the form

$$L_T = -\frac{NT}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\left(\log\det\hat{\boldsymbol{H}}_t + \log\det\boldsymbol{C} + \boldsymbol{u}_t'\hat{\boldsymbol{H}}_t^{-1/2}\boldsymbol{C}^{-1}\hat{\boldsymbol{H}}_t^{-1/2}\boldsymbol{u}_t\right) \tag{10.19}$$

where $\hat{\boldsymbol{H}}_t = \text{diag}(\boldsymbol{h}_t)$ and $\boldsymbol{C}$ is the $N\times N$ contemporaneous correlation matrix, having unit diagonal,. In this set-up, the off-diagonal elements of $\boldsymbol{C}$ are additional parameters to be estimated, and are constrained by a logistic mapping to lie in $(-1, +1)$.

If the equations are simultaneous, with $\boldsymbol{B} \neq \boldsymbol{I}$ in (5.1), the term $T\log|\det\boldsymbol{B}|$ is added to the maximand in (10.19) to implement FIML.

10.2.4 Student-*t* ML with Conditional Heteroscedasticity

To estimate a system with multivariate Student's t errors, the maximand is

©James Davidson 2015

$$L_T = T\log\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} - \frac{T}{2}\log\pi(\nu-2)$$
$$-\frac{1}{2}\sum_{t=1}^{T}\left(\log\det\hat{\boldsymbol{H}}_t + \log\det\boldsymbol{C} + (\nu+1)\log\left(1+\frac{\boldsymbol{u}_t'\hat{\boldsymbol{H}}_t^{-1/2}\boldsymbol{C}^{-1}\hat{\boldsymbol{H}}_t^{-1/2}\boldsymbol{u}_t}{(\nu-2)}\right)\right) \tag{10.20}$$

Note that in this set-up, the degrees of freedom are constrained to be equal for all equations.

10.2.5 GED ML with Conditional Heteroscedasticity

A system with GED errors has maximand

$$L_T = T\log\frac{\nu 2^{(1+1/\nu)}}{\lambda_\nu\Gamma(1/\nu)} - \frac{1}{2}\sum_{t=1}^{T}\left(\log\det\boldsymbol{H}_t + \log\det\boldsymbol{C} + \left|\frac{\boldsymbol{u}_t'\hat{\boldsymbol{H}}_t^{-1/2}\boldsymbol{C}^{-1}\hat{\boldsymbol{H}}_t^{-1/2}\boldsymbol{u}_t}{\lambda_\nu^2}\right|^{\nu/2}\right) \tag{10.21}$$

If the equations are simultaneous, with $\boldsymbol{B} \neq \boldsymbol{I}$ in (5.1), the term $T\log|\det\boldsymbol{B}|$ is added to the maximands in (10.19), (10.20) or (10.21) to implement FIML.

System versions of the skewed-Student and Whittle likelihoods are not implemented.

10.3 Markov Switching Models

In switching-regime models, the likelihood functions take the form of (8.4) where the regime probabilities are generated by the recursive formulae (8.2) and (8.3), and

$$f(Y_t \mid S_t = j, \boldsymbol{\Psi}_{t-1}) = \exp\{l_{jt}\} \tag{10.22}$$

where l_{jt} is the relevant likelihood contribution; in other words, the tth term from the relevant case of L_T, as defined above, evaluated at the parameters for regime j. This can be viewed as an application of the EM algorithm (Dempster, Laird and Rubin 1977) in which the latent indicators of the different regimes, $I(S_t = j)$ are replaced by their conditional expected values, i.e., the conditional probabilities of regime j prevailing at dates t.

©James Davidson 2015

11 Standard Errors and Covariance Matrix Formulae

Four alternative estimators are available for the standard errors, and covariance matrices used to compute the standard test statistics: $\boldsymbol{V}_1$ = 'Standard', $\boldsymbol{V}_2$ = 'Robust' and $\boldsymbol{V}_3$ = 'HAC', $\boldsymbol{V}_4$ = 'KVB'. Let $\boldsymbol{Q}$ denote an estimate of the expected Hessian matrix of the criterion function, and $\boldsymbol{A}$ an estimate of the covariance matrix of the gradient, $\partial C/\partial\boldsymbol{\theta}$. For the log-likelihood and least squares criteria $\boldsymbol{Q}$ is the actual numerical Hessian of C_T at the estimated point. For GMM $\boldsymbol{Q}$ is computed, in effect, by replacing u_t by $\partial u_t/\partial\boldsymbol{\theta}$ (and its transpose) in the formula for C_T. The general formulae then take the form $\boldsymbol{V}_i = \boldsymbol{Q}^{-1}\boldsymbol{A}_i\boldsymbol{Q}^{-1}$, for i = 1,2,3.

11.1 Information Matrix Formulae

For the log-likelihood cases, $\boldsymbol{A}_1 = -\boldsymbol{Q}$ resulting in the information matrix estimate of the covariance matrix. For least squares and GMM, $\boldsymbol{A}_1 = s^2\boldsymbol{Q}$ where s^2 is the residual variance. These standard formulae are not robust, depending on the assumption that the likelihood function is correctly specified (not a quasi-likelihood) and that the residuals are i.i.d. They are recommended only for comparability with other packages.

11.2 Robust Formulae

In the robust formulae, recommended for general use, $\boldsymbol{A}_2$ is set to the outer product of the gradient (OPG). In the MLE and OLS cases, this has the form

$$\boldsymbol{A}_2 = \sum_{t=1}^{T} \frac{\partial c_t}{\partial\boldsymbol{\theta}} . \frac{\partial c_t}{\partial\boldsymbol{\theta}'} \tag{11.1}$$

where c_t is the tth term of the log-likelihood or sum of squares. In the linear least squares case this corresponds to the White (1980) formula. In GMM, $\dfrac{\partial c_t}{\partial\boldsymbol{\theta}}$ is replaced by $\left(\sum_{t=1}^{T} \dfrac{\partial u_t}{\partial\boldsymbol{\theta}} \boldsymbol{z}'\right)\hat{\boldsymbol{M}}^{-1}\boldsymbol{z}_t u_t$. (NB these terms sum to the gradient $\partial C_T/\partial\boldsymbol{\theta}$ in each case.)

11.3 HAC Variance Estimators

The heteroscedasticity and autocorrelation consistent formulae replace the OPG covariance estimate with the Newey-West (1987) type of estimate, using a kernel function to taper the contribution of serial covariances. This formula is recommended when there is evidence of residual serial dependence in the equation. There is a choice of Parzen, Bartlett, Quadratic-Spectral and Tukey-Hanning kernels, with Parzen the default.

Note: The same options are available for the Whittle estimator, although in this case the terms in the OPG and HAC formulae represent coordinates in the frequency domain, not the time domain. Application in this context is at the user's discretion.

11.3.1 Bandwidth Selection

The bandwidth can be set manually, or by an automatic "plug-in" procedure. In the latter case, the bandwidth is set as [m], the largest integer below

©James Davidson 2015

$$m = \begin{cases} 1.1447(\alpha T)^{1/3}, & \text{Bartlett kernel} \\ 2.6614(\alpha T)^{1/5}, & \text{Parzen kernel} \\ 1.3221(\alpha T)^{1/5}, & \text{Quadratic spectral kernel} \\ 1.7462(\alpha T)^{1/5}, & \text{Tukey-Hanning kernel} \end{cases} \tag{11.2}$$

where α is a data-dependent factor. Let the vector of variables whose covariance matrix is to be computed (generally the criterion gradients) be denoted $(x_{1t},\ldots,x_{pt})' = \boldsymbol{x}_t$. Following Andrews (1991) and Andrews and Monahan (1992), α can be chosen as

$$\alpha_A = \begin{cases} \dfrac{\sum_{j=1}^{p} w_j \dfrac{4\hat{\rho}_j^2\hat{\sigma}_j^4}{(1-\hat{\rho}_j)^6(1+\hat{\rho}_j)^2}}{\sum_{j=1}^{p} w_j \dfrac{\hat{\sigma}_j^4}{(1-\hat{\rho}_j)^4}}, & \text{Bartlett kernel} \\ \dfrac{\sum_{j=1}^{p} w_j \dfrac{4\hat{\rho}_j^2\hat{\sigma}_j^4}{(1-\hat{\rho}_j)^8}}{\sum_{j=1}^{p} w_j \dfrac{\hat{\sigma}_j^4}{(1-\hat{\rho}_j)^4}}, & \text{other cases} \end{cases} \tag{11.3}$$

where $\hat{\rho}_j$ is the first-order autoregression coefficient of the x_{jt} variable, and $\hat{\sigma}_j^2$ the corresponding residual variance. TSM chooses the weight w_j to be the reciprocal of the sample variance of x_{jt}.

Following Newey and West (1994), α can also be chosen as

$$\alpha_{NW} = \frac{\left(2\sum_{j=1}^{[n_T]} j\hat{\gamma}_j\right)^2}{\left(\hat{\gamma}_0 + 2\sum_{j=1}^{[n_T]}\hat{\gamma}_j\right)^2} \tag{11.4}$$

where $\hat{\gamma}_j$ is the jth order autocovariance of the series $\boldsymbol{w}_t'\boldsymbol{x}_t$ and $\boldsymbol{w}_t = (w_{1t},\ldots,w_{pt})'$ is the same weight vector as in α_A, and

$$n_T = \begin{cases} 3(T/100)^{2/9}, & \text{Bartlett kernel} \\ 4(T/100)^{2/25}, & \text{Quadratic Spectral kernel} \\ 4(T/100)^{4/25}, & \text{Parzen and Tukey-Hanning kernels} \end{cases} \tag{11.5}$$

See also Den Haan and Levin (1997) for all formulae, details and recommendations.

11.3.2 Pre-whitening

The pre-whitening option computes a VAR(1) model

$$\boldsymbol{u}_t = \boldsymbol{A}\boldsymbol{u}_{t-1} + \boldsymbol{\varepsilon}_t \tag{11.6}$$

for the series $\boldsymbol{u}_t$ (for example, the scores in parameter covariance matrix estimation) and applies the kernel estimator to the residuals from this regression. Call this estimate $\boldsymbol{U}_3$. This matrix is 're-coloured', to yield the HAC covariance matrix for $\boldsymbol{u}_t$ as

©James Davidson 2015

$$\boldsymbol{V}_3 = (\boldsymbol{I} - \hat{\boldsymbol{A}})^{-1}\boldsymbol{U}_3(\boldsymbol{I} - \hat{\boldsymbol{A}}')^{-1}. \tag{11.7}$$

Use the pre-whitening option with caution, since its properties in particular cases are unclear. If the scores are highly collinear, the VAR estimate could be poorly conditioned and $\boldsymbol{I} - \hat{\boldsymbol{A}}$ close to singular. Over-compensating for autocorrelation (by pre-whitening and/or using plug-in bandwidths) could effectively kill the power of the tests of I(0).

11.4 KVB Inconsistent Variance Estimates

A fourth option implemented is that of Kiefer, Vogelsang and Bunzel (2000) (KVB). These authors suggest using an inconsistent estimator of the covariance matrix, which yields asymptotically pivotal statistics having a non-standard distribution. As shown in Kiefer and Vogelsang (2002a, 2002b), an equivalent procedure is obtained by using the HAC estimator with the Bartlett kernel and bandwidth set equal to sample size.

p-value inequalities are reported using the tabulation in Kiefer and Vogelsang (2002b). See the cited papers for details. The method is also implemented for nonlinear models/hypotheses, See Bunzel, Kiefer and Vogelsang (2001) on these applications.

©James Davidson 2015

12 Test Statistics

Three types of test statistic can be computed: Wald tests, Lagrange multiplier (LM) tests, also known as score tests, and moment tests (M tests). For the background theory and literature references on these tests, see Davidson (2000) Chapter 12. Under standard regularity conditions, all these tests have asymptotic $\chi^2(r)$ distributions when the null hypothesis is true, with the number of degrees of freedom r corresponding to the number of restrictions under test.

12.1 Wald Tests

Wald tests are computed using the formula

$$\mathrm{W} = \hat{\boldsymbol{g}}'(\hat{\boldsymbol{G}}\hat{\boldsymbol{V}}\hat{\boldsymbol{G}}')^{-1}\hat{\boldsymbol{g}} \tag{12.1}$$

where $\boldsymbol{g}$ denotes a vector of r linear or nonlinear restrictions, and $\boldsymbol{G} = \partial\boldsymbol{g}/\partial\boldsymbol{\theta}'$ ($r \times p$) is the Jacobian matrix of the restrictions. In many cases this is just a selection matrix picking out a subset of the parameters, to test that they are jointly equal to zero. The 'hats' denote evaluation at unrestricted parameter values. Nonlinear restrictions are implemented by typing the specified formulae in standard notation (see Section 4.6.1) and evaluating the Jacobian numerically.

Ordinary t-values (ratios of estimates to standard errors) are asymptotically normal under the null hypothesis, and can be thought of as Wald statistics ($\chi^2(1)$) after squaring. These statistics are not given explicitly, but the nominal p-values for 2-sided tests are quoted where appropriate, as for all the quoted statistics.

12.2 Lagrange Multiplier Tests

The basic statistic is obtained from the formula

$$\mathrm{LM} = \dot{\boldsymbol{q}}'\dot{\boldsymbol{Q}}^{-1}\dot{\boldsymbol{G}}'(\dot{\boldsymbol{G}}\dot{\boldsymbol{V}}\dot{\boldsymbol{G}}')^{-1}\dot{\boldsymbol{G}}\dot{\boldsymbol{Q}}^{-1}\dot{\boldsymbol{q}} \tag{12.2}$$

where $\boldsymbol{q}$ denotes the gradient of the unrestricted model criterion and the 'dots' denote evaluation at the restricted estimates. The matrix $\boldsymbol{V}$ in (12.1) and (12.2) is computed by the currently selected formula, either simple, robust or HAC. Being based on the restricted model, these tests typically have a diagnostic role.

12.3 Moment and Conditional Moment Tests

Moment tests are tests of hypotheses of the form $E(\boldsymbol{m}) = \mathbf{0}$ where $\boldsymbol{m}$ is a vector of sample moments of functions depending on the model parameters. The leading cases are covariances of model disturbances, or squared model disturbances, with a set of test variables. In general these variables are indicators of incorrect specification, so that these tests also have a diagnostic role. The test statistic takes the form

$$\mathrm{M} = \dot{\boldsymbol{m}}'\dot{\boldsymbol{V}}_m^{-1}\dot{\boldsymbol{m}} \tag{12.3}$$

where $\boldsymbol{V}_m$ is the asymptotic covariance matrix of the moments, taking into account their dependence on unknown parameters $\boldsymbol{\theta}$, and all terms are evaluated under the null hypothesis that the estimated model is correct. If $\boldsymbol{V}_m$ is computed using a formula that assumes the terms of $\boldsymbol{m}$ are uncorrelated, the null hypothesis can be thought of as taking the form $E(\boldsymbol{w}_t|\Psi_{t-1}) = \mathbf{0}$, $t = 1, \ldots, n$, where $\boldsymbol{m} = n^{-1}\Sigma\boldsymbol{w}_t$, $\boldsymbol{w}_t$ is a chosen function of data for observation t and parameters, and Ψ_t denotes the history of the process to date. In this form

©James Davidson 2015

they are called conditional moment (CM) tests. To test the less restrictive simple moment hypothesis (that the unconditional mean of the terms is zero) $\boldsymbol{V}_m$ should be computed using an HAC formulation, in case the data are not serially independent.

LM and CM tests can be computed both for user-selected hypotheses and a range of pre-set diagnostic tests. Table 1 shows the cases available in the latter tests. Note that the LM test is implemented by computing the scores of the criterion functions numerically, after extending the specification of the model as indicated. Therefore it does not depend on extra simplifying assumptions, such as normal disturbances.

Table 1. Dummy Alternative Hypotheses in Diagnostic Tests

	LM Test	CM Test
A. Residual Autocorrelation	Lagged residuals in equation (4.1) or equation (5.1)	Covariances of current and lagged normalized* residuals
B. Neglected ARCH	Lagged squared residuals in equation (6.1) or equation (7.1)**	Covariances of current and lagged squared normalized* residuals
C. Nonlinear functional form (RESET)	Integer powers of the fitted values in equation (4.1) or equation (5.1).	Covariances of normalized* residuals with integer powers of fitted values (mean deviations).
D. Heteroscedasticity	Square of fitted values in equation (6.1) or equation (7.1)**	Covariances of squared normalized* residuals with squared fitted values.
E. White's Heteroscedasticity Test***	Squares and products of explanatory variables in equation (6.1) or equation (7.1).**	Covariances of squared normalized* residuals with squares and products of explanatory variables.
F. AR Common Factors****	p lags of variables x_{1t} in equation (4.1) or equation (5.1) included in x_{2t}	Covariances of normalized residuals with p lags of x_{1t}

* In the case of ARCH-type or Markov-switching models, "normalized" means that the residuals have been divided by their estimated conditional SDs. Their squares, after subtracting 1, are *iid*(0,1) by the null hypothesis.

** In the case where no conditional variance model is specified under H_0, the statistic is computed from the regression of the squared residuals on the test variables. Otherwise, the test variables are added to the CV model.

*** In case the indicators in the White tests are linearly dependent, the statistic is computed with the maximum available principle components, with degrees of freedom adjusted accordingly.

**** The common factor tests are available in the case where under H_0 there are no lags on x_{1t}, and no variables of Type 2 are specified.

In systems of equations, the diagnostic tests include the test variables (e,g, lagged residuals) from every equation in every equation. Hence for example, the number of degrees of freedom for an autocorrelation test with one lag in a system of two equations is four.

©James Davidson 2015

The LM and CM tests of corresponding hypotheses can give different results, since they are based on different assumptions. Thus, CM tests have the advantage that they impose minimal assumptions about the form of the model under the alternative hypothesis, but they can require the existence of higher-order moments for validity.

12.4 Information Matrix Test

This is a diagnostic for maximum likelihood estimation, with a CM-type motivation. The test, in effect, compares the matrices $\boldsymbol{A}_2$ in (11.1) and $\boldsymbol{Q}$ term by term, with the null hypothesis that the expected difference is zero. Thus, letting $c_t = -l_t$ where l_t denotes the t^{th} contribution to the log-likelihood, and defining

$$\boldsymbol{q}_t = \frac{\partial c_t}{\partial \theta}, \qquad \boldsymbol{Q}_t = \frac{\partial^2 c_t}{\partial \theta \partial \theta'}, \tag{12.4}$$

the moments under test are of the form

$$\boldsymbol{m} = \mathrm{Vec}\left(n^{-1} \sum_{t=1}^{n} \left(\hat{\boldsymbol{Q}}_t - \hat{\boldsymbol{q}}_t \hat{\boldsymbol{q}}_t' \right) \right)^* \tag{12.5}$$

where the hats denote evaluation at estimated values, and the * denotes the omission of elements redundant through symmetry. See Davidson (2000) Section 12.5.4 for further details. If either this test or the heteroscedasticity tests reject, the use of the robust covariance matrix option is strongly advised.

12.5 Nyblom-Hansen Stability Tests

These are useful general-purpose tests for the stability of an estimated model, based on the partial sums of the gradient contributions; see Nyblom (1989), Hansen (1990, 1992). They are valid for any of the time domain estimators implemented in TSM4, and can be computed for the model as a whole and also, optionally, for the parameters of the model one at a time. Let

$$\boldsymbol{s}_t = \sum_{j=1}^{t} \boldsymbol{q}_j \quad (p \times 1) \tag{12.6}$$

where $\boldsymbol{q}_j$ is defined in (12.4). The full model test statistic (called the L_C statistic in Hansen 1990) is then

$$\mathrm{NH}(p) = \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{s}_t' \left(\sum_{t=1}^{n} \boldsymbol{q}_t \boldsymbol{q}_t' \right)^{-1} \boldsymbol{s}_t \tag{12.7}$$

For testing parameters for constancy individually, the statistics have the form

$$\mathrm{NH}(1) = \frac{1}{n} \left(\sum_{t=1}^{n} q_{jt}^2 \right)^{-1} \sum_{t=1}^{n} s_{jt}^2 \,, j = 1, \ldots, p \tag{12.8}$$

Note that $\hat{\boldsymbol{s}}_n = \boldsymbol{0}$ by construction. When the model is correctly specified, the normalized partial sum process $\{\boldsymbol{s}_t\}$ should behave like the mean deviations of a martingale, and hence in the limit like a Brownian bridge. Rejections are expected when the elements exhibit features such as shifts of mean or other time-dependent patterns. While set up as tests for random parameter variation, the NH tests should hopefully have power against a wide range of misspecifications.

©James Davidson 2015

12.6 Andrews Structural Change LM Test

This is the LM version of the test developed in Andrews (1993) to detect structural change with an unknown change point. In Andrews' paper the test is defined for changes in any subset of the model parameters. In this implementation, the options are to test the full parameter set, and to test each parameter individually. This test is implemented for models estimated by least squares, maximum likelihood and GMM. The test formula is $\sup_{\pi\in[\pi_1,\pi_2]} LM(\pi)$ where

$$LM(\pi) = \frac{T}{\pi(1-\pi)} \bar{\boldsymbol{m}}_1(\pi)' \hat{\boldsymbol{S}}^{-1} \hat{\boldsymbol{M}} \left(\hat{\boldsymbol{M}}' \hat{\boldsymbol{S}}^{-1} \hat{\boldsymbol{M}} \right)^{-1} \hat{\boldsymbol{M}}' \hat{\boldsymbol{S}}^{-1} \bar{\boldsymbol{m}}_1(\pi) \tag{12.9}$$

Consider first the least squares/maximum likelihood implementation of the test. In (12.9), $\bar{\boldsymbol{m}}_1(\pi) = \sum_{t=1}^{T\pi} \hat{\boldsymbol{m}}_{1t}$ where the $\hat{\boldsymbol{m}}_{1t}$ are the score (gradient) contributions corresponding to the parameters under test, evaluated at the estimates under the null hypothesis and hence having the property $\bar{\boldsymbol{m}}_1(1) = \mathbf{0}$. In the full-model stability test $\boldsymbol{m}_1$ is the complete score vector. In the individual parameter tests, it is a scalar, the corresponding score element. $\boldsymbol{S}$ is the covariance matrix of the (full) score vector, and the estimated matrix $\hat{\boldsymbol{S}}$ may be computed by the robust or HAC formulae, according the program settings selected. $\boldsymbol{M}$ is the matrix of derivatives of $\boldsymbol{m}_1$, or in other words the log-likelihood Hessian, or columns thereof in the case of individual parameter tests. Note that in the full-model test, $\boldsymbol{M}$ is square $p \times p$ and the formula in (12.9) reduces to

$$LM(\pi) = \frac{T}{\pi(1-\pi)} \bar{\boldsymbol{m}}_1(\pi)' \hat{\boldsymbol{S}}^{-1} \bar{\boldsymbol{m}}_1(\pi) . \tag{12.10}$$

In the GMM implementation, typically $\bar{\boldsymbol{m}}_1(\pi) = \sum_{t=1}^{T\pi} z_t \hat{u}_t$ $(m \times 1)$ where $m \geq p$ the column dimension of the instrument matrix $\boldsymbol{Z}$, $\boldsymbol{S} = T^{-1}\boldsymbol{Z}'\boldsymbol{Z}$ and $\boldsymbol{M} = \boldsymbol{Z}'\boldsymbol{G}$ where $\boldsymbol{G}$ is the matrix whose rows are the derivatives of u_t with respect to the test parameters. Note that in this case, the orthogonality condition imposed by estimation takes the form $\hat{\boldsymbol{M}}' \hat{\boldsymbol{S}}^{-1} \bar{\boldsymbol{m}}_1(1) = \mathbf{0}$.

The test is implemented by maximizing the expression in (12.9) by a grid search over the interval $[\pi_1, \pi_2]$ for $0 < \pi_1 < \pi_2 < 1$. The bounds are user-selectable with default values of $\pi_1 = 0.1$, $\pi_2 = 0.9$. Andrews' (1993) table of simulated critical values is used to compute *p*-value bounds.

12.7 Consistent Specification Tests

Note: the tests in this section are advanced and, to simplify the program interface, dialog elements to control them are not displayed by default. See the user's manual Section 8.2a for further information.

12.7.1 Bierens Tests

This is one of a class of consistent CM tests of functional form proposed by Bierens (1990). The test covariance is between the model residuals and a bounded nonlinear function of the explanatory variables. Let $\boldsymbol{x}_t$ denote the vector of all exogenous variables in the mean model. The test function takes the form

$$w_t(\xi) = \exp(\xi' \arctan(\tilde{\boldsymbol{x}}_t)) \tag{12.11}$$

©James Davidson 2015

where $\tilde{\boldsymbol{x}}_t$ denotes the standardized series, with elements expressed in deviations from means and divided by standard deviations, and the vector $\xi\,(K\times 1)$ has elements drawn from the interval $[-1,1]$.

Setting ξ is to a fixed value ξ_0, the corresponding statistic S_0 is asymptotically $\chi^2(1)$ under the null hypothesis of correct specification, as for any fixed choice of ξ, but this initial choice is not necessarily optimal. To remove dependence of the test on ξ, consider an integrated statistic computed by one of the following formulae, where Ξ is a compact subset of $\mathbb{R}^K$:

$$\hat{S}^A = \int_\Xi S(\xi)d\xi \tag{12.12}$$

$$\hat{S}^B = 2\log\int_\Xi \exp\left\{\tfrac{1}{2}S(\xi)\right\}d\xi \tag{12.13}$$

$$\hat{S}^S = \sup_{\xi\in\Xi} S(\xi)\,. \tag{12.14}$$

See Andrews and Ploberger (1994) for an analysis of tests of this type. These integrated statistics have unknown null distributions, and two approaches exist to computing *p*-value. The first is the bootstrap method described by Hansen (1996), which can, however, be quite computationally intensive. Alternatively, Bierens (1990) suggests defining the statistic

$$\tilde{S} = \begin{cases} S_0, \hat{S} - S_0 \le \gamma T^\rho \\ \hat{S}, \hat{S} - S_0 > \gamma T^\rho \end{cases} \tag{12.15}$$

where $\gamma > 0$ and $0 < \rho < 1$. Judicious choice of the bound ensures that $\tilde{S} = S_0$ with probability approaching 1 when H_0 is true, and hence the test based on χ^2 critical values is asymptotically correctly sized. Under the alternative, note that $\tilde{S}$ must diverge at the rate of $\max(\hat{S}, S_0)$, hence ensuring the power of the optimized test.

In this implementation, the bound in (12.15) is computed as

$$Bound = \gamma_0 T^{\rho_1} D^{\rho_2}(1 + I(\text{sup test}, K > 1)K^{\rho_3}) \tag{12.16}$$

where $I(\bullet)$ is the indicator function, "sup test" means that option (12.14) is selected and γ_0, ρ_1, ρ_2 and ρ_3 are user-selectable. The set Ξ is constructed as a *K*-dimensional hypercube whose sides have user-selectable upper and lower bounds. However, ξ is normalized to fix the range of variation of $\xi'\arctan(\tilde{\boldsymbol{x}}_t)$ over the sample, also user-selectable. Hence, there is little reason to change these from the defaults, other than to introduce asymmetry around zero.

12.7.2 Score-contribution Tests

These tests generalize the Bierens testing procedure by using score-contributions in place of residuals as the target series to be tested for correlation with (12.11). Either a *p*-degree of freedom test can be constructed using the complete score vector, or 1-degree of freedom tests based on individual score elements.

12.7.3 Dynamic Specification Tests

The Bierens and score-contribution tests can be considered exclusively as tests of functional form, in which case $\boldsymbol{x}_t$ in (12.11) can be thought of as variables appearing in the

©James Davidson 2015

model only. These tests can also test for omitted variables and especially dynamic specification, in which case $\boldsymbol{x}_t$ can include variables not in the null model. In dynamic models, testing for omission of lags poses the problem of specifying a lag truncation point. An option in these tests is to include an indefinite number of lags (increasing with sample size) while keeping the number of nuisance parameters fixed, by means of a polynomial distributed lag scheme. Thus, define the test function as

$$w_t(\xi) = \exp\left\{\sum_{j=0}^{t-1} \lambda_j' \arctan(\tilde{r}_{t-j})\right\} \tag{12.17}$$

where r_t $(m\times 1)$ is the vector of test variables and

$$\lambda_j = \sum_{i=1}^{P} (j+1)^{-i-1}\xi_i \,. \tag{12.18}$$

The nuisance parameters are $\xi = (\xi_1', \ldots, \xi_P')'$ of dimension $K = Pm$. Even a modest choice of *P* can model a large range of lag distributions. See Davidson and Halunga (2012) for further details, and some Monte Carlo evidence.

12.8 Vuong's Test of Non-Nested Models

Vuong (1989) describes test procedures for comparing non-nested models. If $\hat{l}_{1t}$ and $\hat{l}_{2t}$ for $t = 1,\ldots,T$ are the log-likelihood contributions for each fitted model, and $m_t = \hat{l}_{1t} - \hat{l}_{2t}$, are their differences, the normalized likelihood ratio statistic

$$V = \sqrt{T}\bar{m} / s_m \tag{12.19}$$

where $\bar{m}$ and s_m are respectively the sample mean and standard deviation of m_t, can be used as a guide to model selection. When the data cannot discriminate between the rival specifications *V* is asymptotically distributed as standard normal under specified regularity conditions, otherwise its sign indicates the ranking of the alternatives.

Such tests cannot be pre-programmed since the models must be estimated individually, but the series of log-likelihood contributions for any model can be retrieved and added to the data set for subsequent analysis. The statistic is then easily computed by hand using TSM's various data-handling features. Use the Data Transformation and Editing dialog to form the differences of two series, then the Compute Summary Statistics dialog for the mean and standard deviation. The Calculator dialog can then compute the statistic and, finally, consult Look Up Tail Probability for an asymptotic *p*-value.

12.9 Cusum of Squares Test

This is a version of the Brown-Durbin-Evans (1975) test for parameter stability, computed from the regression residuals and corrected for possible autocorrelation and conditional heteroscedasticity using a HAC variance estimator, as derived by Deng and Perron (2008). The statistic has the form

$$CUSQ = \max_{1\le n\le T} \frac{\left|\sum_{t=1}^{n}\left(\hat{u}_t^2 - T^{-1}\sum_{t=1}^{T}\hat{u}_t^2\right)\right|}{\sqrt{T\hat{\varphi}}} \tag{12.20}$$

where $\hat{\varphi}$ is any of the variance estimators described in Section 11.3, computed for the sample mean deviations of $\hat{u}_t^2$. The asymptotic distribution, under the assumptions specified in Deng and Perron (2008), is the supremum of the absolute value of a Brownian

©James Davidson 2015

bridge.[4] The test has power against the alternative of unconditional changes in the variance of the residuals in a regression model. Note that the assumptions require the existence of moments greater than fourth order, so the test may be inappropriate for fat-tailed distributions.

12.10 'Sup-F' Tests

A number of testing situations face the problem that one or more parameters are unidentified under the null hypothesis. Say, a statistic F depends on a parameter π which cannot be estimated under H_0. A solution to this problem is to base the test on the statistic $\sup_{\pi\in\Pi} F(\pi)$ (see e.g. Davies, 1977). The Andrews (1993) LM test is one such case already implemented. However, it is not difficult to compute other statistics of this type in TSM, where π is either one- or two-dimensional, using the Criterion Plot feature. This will evaluate the model, and any specified test statistics, over a specified grid of fixed values of one or two model parameters. The output reports the extreme values attained over the grid for all specified test statistics, including absolute *t*-ratios, diagnostic statistics, user-specified Wald tests, etc. There is no limit apart from computing time to the number of grid points that can be evaluated. For applications, see the sections above on regime-switching and date-transition models.

12.11 Tests of Integration Order

A range of tests on individual time series of the parameter d in the representation I(d) can be computed through the Setup / Compute Summary Statistics dialog. To compute these tests after allowing for a linear trend in the series, check the "Detrend" option in the dialog. In the cases of the nonparametric tests that employ a kernel estimate of the long run variance, The kernel and bandwidth settings are selectable in the Options / Tests and Diagnostics dialog.

12.11.1 Tests of I(0)

Robinson-Lobato Test This is the signed, one-dimensional version of the nonparametric test based on the periodogram proposed in Robinson and Lobato (1998). The statistic is $-t$ where t is defined in (2.2) that paper, which is standard normal under the null hypothesis. This provides a test against the alternative $d > 0$ with rejection in the upper tail, and also a test against the alternative $d < 0$ with rejection in the lower tail.[5] The *p*-value quoted is for the former test. The bandwidth m is chosen using the formula advocated in Section 3 of that paper.

KPSS Test. This is the test is due to Kwiatkowski et al. (1992). The quoted *p*-value inequalities use the table given in the cited paper.

V/S Test Modified version of the KPSS test due to Giraitis et al (2003). The *p*-values are computed analytically using the formula given in the paper.

Lo's modified R/S Test (Lo 1991). This the version of Hurst's test for short memory using a kernel HAC estimator of the variance. The quoted *p*-value inequalities use the table given in the cited paper.

---

[4] This is the same as the distribution of the Kolmogorov-Smirnov test for the equality of empirical distributions, note.

[5] Note that these authors assign the tails of their statistic incorrectly in their discussion.

©James Davidson 2015

Harris-McCabe-Leybourne Test (Harris, McCabe and Leybourne 2008). The HML test is based on the long-range autocovariances, and is asymptotically N(0,1) under the null hypothesis of short memory. Two settings need to be selected, a truncation parameter c where $k = (cT)^{1/2}$ is the lowest-order lag, and L, where $l = LT^{12/25}$ is the bandwidth truncation of the variance parameter. See the Harris et al. paper for details. These settings can be changed in the Options / General dialog under "Special Settings".

- All of these tests, except the Robinson-Lobato test, can be computed post-estimation for model residuals as well as for raw data. See the Diagnostics dialog under Options / Tests and Diagnostics to select these options. In this case, note that the HML bias correction is implemented for nonlinear as well as linear models, by substituting derivatives for measured linear regressors. This extension is natural, but is not dealt with by HML in their paper, hence it should be used with caution.

12.11.2 Tests of I(1)

Augmented Dickey-Fuller Test (Dickey and Fuller 1979, Said and Dickey 1984). The number of lags for the ADF test is chosen from $[0,\ldots, 2T^{1/3}]$ by a user-selectable information criterion - see Options / Tests and Diagnostics, "Model Selection Criterion". If the option is set to "None", the lag length must be set manually using a scroll bar in the same dialog.

Phillips-Perron Test (Phillips and Perron 1988) .

Elliott-Rothenberg-Stock Tests (Elliott et al. 1996). There are two tests proposed in this paper, a modified Dickey-Fuller test (DF-GLS) and a feasible likelihood ratio test which these authors call the P_T test. The former chooses lags using and information criterion, as for the regular ADF, and the latter uses a kernel estimate of the variance which is computed as for the other tests in this dialog.

- The ADF and PP tests of I(1) can also be computed in the Model / Linear Regression dialog, by running a regression of the test series on intercept and optionally trend. In this case, the lag length for the ADF test can be chosen manually by the user.
- The critical values for these tests are taken from the tables reported in Elliott et al (1996) and Fuller (1976) respectively. The entries in these tables for finite sample sizes are linearly interpolated to provide a value appropriate to the actual sample being analysed, and for this purpose the case "∞" is treated as equivalent to 1000 observations.

12.12 Bootstrap Test of I(0)

This test is proposed in Davidson (2009). Defining "I(0)" to be the property that the normalized partial sums of a time series converge to Brownian motion, and hence yield the standard asymptotic distributions postulated by cointegration theory, the object of this test is the determine whether the approximation to such distributions is 'adequate' in a sample of given size. This is done by simulating the series using a fitted model, and computing the statistic defined by Breitung (2002) for the partial sums. The same statistic is also computed for an i.i.d. sequence, which is to say, the shocks used to drive the model simulation – these can be either computer generated, or randomly resampled residuals. These bootstrap distributions are compared using the Kolmogorov-Smirnov test for equality of empirical distributions. The reported p-values are obtained from the formula in Feller (1948).

©James Davidson 2015

Observe the following points.

1. The test is motivated as an attempt to answer the question "Will asymptotic distribution results based on the assumption of I(0) provide more accurate approximate inferences than alternatives, in my sample?" Test outcomes depend on sample size as well as the form of the data generation process (DGP). It provides an alternative to tests such as the KPSS whose performance tends to be excessively dependent on the choice of bandwidth.
2. Don't overlook that the test is performed on the fitted model of the series, not the series itself. Test performance depends on how well the DGP is able to reproduce the autocorrelation characteristics of the data to which it is fitted. Careful fitting and experimentation with different models (for example, ARMA and ARFIMA variants) is recommended. The choice of shock distribution (Student's *t* versus Gaussian, for example) may also affect the outcome.

Two variants of the I(0) test are implemented. The one most easily performed is to model the dependence of the series automatically by a sieve autoregression, choosing lag length by the Akaike criterion with maximum lag length limited to $O(T^{1/3})$. This is fast, but needs a linear representation of the series to be adequate.

The second variant uses a model constructed by the investigator to represent the dependence, and hence can include such features as long memory and nonlinearities as well as permitting more care in validating the model specification. In this case, the i.i.d. sequence is generated from the investigator's model in which all time-dependence parameters are suppressed, but disturbance distribution parameters are retained. This variant is valid, in particular, for dynamic binary and count data models.

©James Davidson 2015

13 Simulation and Resampling Options

A special feature of the program is the simulation capability. Any model that can be estimated can also be simulated. A range of methods for generating the random shocks is provided, including the bootstrap, and model-specific distributions including the Gaussian, Student *t*, GED, as well as probit, logit and Poisson for the discrete data cases. Simulations can be run on a "one-off" basis, and comparing the appearance of series generated from a fitted model with the original data is an excellent informal method of specification checking.

13.1 Monte Carlo Experiments

The Monte Carlo option provides the means to study the distributions of estimators and test statistics. The reported statistics from the experiment can include the following:

1. Moments of the distributions of parameter estimates (mean, variance, skewness, kurtosis).
2. Bias and RMSE estimates. For this option the parameter sets of the simulated and estimated models must match, but restrictions can be placed on the estimated model to simulate misspecification.
3. Upper-tail quantiles of all test statistics, including parameter absolute or signed *t* values, diagnostic statistics and additional Wald, LM and moment tests. These outputs can be used to generate tables of critical values for tests for given sample sizes that may improve on the asymptotic approximations.
4. Empirical distribution functions of *p*-values. These should be uniformly distributed when the null is true and the tests are correctly sized. These outputs are useful to assess the size and power of tests under different data generating setups.

To run an experiment, a model to generate the data (DGM) must be specified and stored using the Model Manager (see the Setup menu). Simply set up the model specification in the usual manner for estimation. Parameter values can be the results of the most recent estimation run, or they can be set manually in the Values dialogs. A shock distribution must be selected in Bootstrap and Simulation Options. The 'bootstrap' options randomly resample the residuals from the latest run. Also, specify a model to be estimated (EM), which can be the same as the DGM, or different. The latter option allows misspecification analysis.

13.2 Parallel Processing

Monte Carlo experiments can be either run interactively or launched as external batch jobs, running from the Windows command line. There is a facility to launch a set of identical parallel runs and later combine their outputs into a single set of results. In this way, dual core or quad-core processors can be exploited to run experiments two or four times faster as with a single processor, while at the same time keeping TSM free for other tasks.

13.3 Numerical Test Distributions

The empirical distribution functions (EDFs) of simulated test statistics can be written to a spreadsheet file. If a statistic is simulated when the DGM represents a case of the null hypothesis (or equivalently, with the 'centred *t*-statistic' option selected) the tabulations can be used to compute *p*-values in subsequent runs. The rejection frequencies in

©James Davidson 2015

simulations of the alternative hypothesis can then estimate the true (size-corrected) test powers.

The stored EDFs can also be used to calculate *p*-values for models estimated from observed data, in the ordinary way. In one context, this method might operate as an alternative implementation of the bootstrap. However, tabulations of non-standard distributions with a general application might also be generated, to extend or improve published tables, or even to implement completely new tests.

An EDF file can contain tabulations for parameter *t* values and for any statistic computed by TSM, or coded by the user (identified by name). It can also contain tabulations for the same statistic in different sample sizes. These composite tables can be created from the results of several simulation experiments using a "Merge EDF Tables" command.

Tables for multiple sample sizes are interpolated to generate an approximate *p*-value relevant to the actual sample in use. Suppose tabulations exist for the sample sizes $T_1 \le T_2 \le \dots T_N$. If the actual sample size is T where $T_j \le T \le T_{j+1}$, then the *p*-value returned is a weighted average of the T_j and T_{j+1} tabulations with weights w and $1 - w$, where

$$w = \left(\frac{T_{j+1} - T}{T_{j+1} - T_j} \right)^2 . \tag{13.1}$$

If either $T < T_1$ or $T > T_N$ then the nearest tabulation is used, un-weighted. Note that the squared weighting coefficient places greater weight on the larger sample size than a simple linear interpolation would do, as appropriate if the distributions in question enjoy convergence to the limit at the rate $T^{1/2}$.

13.4 Bootstrap Inference

The program optionally computes bootstrap standard errors and *p*-values for test statistics, including *t*-tests on parameters, diagnostic tests and specified tests of model restrictions. See for example Horowitz (2000), Li and Maddala (1996) for background information on the bootstrap.

The main method adopted is the *parametric bootstrap*. In other words, for each bootstrap replication an artificial sample is created by the simulation module. The object is to reproduce the distribution of the test statistic under the null hypothesis. For those tests where estimates under the alternative are used, such as *t* tests and Wald tests, the tests statistic is centred on the 'pseudo-true' parameter values. Thus, the null distribution for the *j*th *t* statistic is computed as the distribution of

$$t^* = \frac{\hat{\theta}_j^* - \hat{\theta}_j}{\text{s.e.}(\hat{\theta}_j^*)} \tag{13.2}$$

where $\hat{\theta}_j$ and $\hat{\theta}_j^*$ are respectively the point estimate from the observed data, and the point estimate from the bootstrap replication. The null distribution of the Wald statistic is generated similarly, by replicating the formula

$$\text{W} = (\hat{\boldsymbol{g}}^* - \hat{\boldsymbol{g}})'(\hat{\boldsymbol{G}}^* \hat{\boldsymbol{V}}^* \hat{\boldsymbol{G}}^{*\prime})^{-1}(\hat{\boldsymbol{g}}^* - \hat{\boldsymbol{g}}) \tag{13.3}$$

where $\hat{\boldsymbol{g}}$ is the vector of restrictions at the sample estimates, and * denotes the value computed from the bootstrap sample. Null distributions for M and CM tests are generated similarly, by expressing the moments under test as deviations from the sample values. In general, bootstrap variants of LM tests are not available, because the null distributions of

©James Davidson 2015

these statistics are cannot be so easily simulated if the null is false in the sample data. However, the important diagnostic tests are tests for the absence of heteroscedasticity and autocorrelation in the disturbances. Since these are i.i.d. by construction in the bootstrap samples, the bootstrap *p*-values should be valid for specification testing, and are quoted. Use them with caution in other contexts. The diagnostic *Q* tests can be interpreted similarly in the bootstrap context, and bootstrap *p*-values are given. However, for obvious reasons, the Jarque-Bera test for normality of the residuals is bootstrapped only in the case where Gaussian disturbances are specified in the resampling.
NOTE: bootstrap *p*-values are always marked with a * in the output for easy recognition.

Bootstrap 95% equal-tail confidence intervals (*L*, *U*) are reported by default. Note that these are not centred on the point estimate. Letting $\hat{\theta}$ denote the estimate from the observed sample, the formulae are respectively

$$L = \hat{\theta} - t_1,\ U = \hat{\theta} + t_2 \tag{13.4}$$

such that 2.5% of the estimates from the bootstrap replications ($\hat{\theta}^*$) lie above $\hat{\theta} + t_1$ and 2.5% lie below $\hat{\theta} - t_2$. To interpret this estimate note that

$$\begin{aligned} P(\hat{\theta} - t_2 \le \hat{\theta}^* \le \hat{\theta} + t_1 \mid \hat{\theta}) &= P(\hat{\theta}^* - t_1 \le \hat{\theta} \le \hat{\theta}^* + t_2 \mid \hat{\theta}) \\ &\approx P(\hat{\theta} - t_1 \le \theta_0 \le \hat{\theta} + t_2) \end{aligned} \tag{13.5}$$

where $P(.\mid\hat{\theta})$ is the probability under the bootstrap distribution estimated by the corresponding proportion of Monte Carlo replications, and θ_0 is the true value. "≈" denotes that the bootstrap distribution is expected to mimic the true sampling distribution to within some approximation.

Alternative confidence intervals are based on the distribution of the bootstrap *t*-ratios t^*, as defined in (13.2). The equal-tailed "percentile-*t*", interval has bounds

$$L = \hat{\theta} - s_\theta^* q_1,\ U = \hat{\theta} + s_\theta^* q_2 \tag{13.6}$$

where q_1 and q_2 are respectively the 0.975 and 0.025 quantiles of the bootstrap distribution of the t^*, and s_θ^* is the standard error of the bootstrap distribution of $\hat{\theta}^*$. The symmetric "percentile-*t*", has bounds

$$L = \hat{\theta} - s_\theta^* q_3,\ U = \hat{\theta} + s_\theta^* q_3 \tag{13.7}$$

where q_3 is the 0.95-quantile of the bootstrap distribution of $|t^*|$. This last interval is centred on the point estimate, unlike the first two, but its tails represent unequal probabilities when the distribution is skewed – otherwise, the intervals should be similar in large samples.

See Hall (1992) Chapter 1 for a good exposition of the bootstrap principle.

Bias correction for point estimates is also optionally implemented. The bias is estimated by $B = E(\hat{\theta}^* \mid \hat{\theta}) - \hat{\theta}$, estimating $E(\hat{\theta}) - \theta_0$, and hence the corrected estimates take the form

$$\tilde{\theta} = \hat{\theta} - B = 2\hat{\theta} - E(\hat{\theta}^* \mid \hat{\theta}) \tag{13.8}$$

In these formulae, $P(.\mid\hat{\theta})$ and $E(.\mid\hat{\theta})$ denote probabilities and expectations under the bootstrap distribution, estimated in practice by Monte Carlo replication.

©James Davidson 2015

The number of replications used to construct the bootstrap distributions is 99 by default. While this number may be increased for more accurate Monte Carlo estimation, increasing it does not increase the accuracy of the bootstrap statistics beyond a certain point, at least in a finite sample, because the error in estimating the parameters (of $O_p(T^{-1/2})$) will start to dominate. 399 is probably sufficient for nearly every purpose.

13.5 Resampling Methods

The following randomization methods are implemented to draw the disturbances for Monte Carlo and bootstrap resampling.

13.5.1 Likelihood Model

The distribution specified for maximum likelihood estimation. This requires the parameters of the model to be specified in the usual manner. Note that in Monte Carlo experiments, it is possible to use one model for data generation and another for estimation, and hence experiment with misspecification.

13.5.2 Gaussian

Independent Gaussian drawings with mean zero. The standard deviation is selectable.

13.5.3 Stable

Independent drawings from a "fat-tailed" stable distribution with infinite variance. The parameters controlling kurtosis (α) and skewness (β) as well as a scale parameter analogous to the standard deviation are selectable

13.5.4 Formula

A user supplied formula (using the program's symbolic algebra capability) to generate any distribution derived from standard Gaussian or uniform[0,1] components, as well as elements of the data and parameter sets. Mixed Gaussian distributions are just one possibility.

13.5.5 Simple/Block Bootstrap

These options resample of the currently stored model residuals. The residuals are centred and multiplied by a variance correction factor of $\sqrt{n/(n-k)}$ where n is sample size and k the number of fitted parameters. The block length L must be set by the user, and to get the simple bootstrap set this to 1. In the simple bootstrap procedure, note the importance of the assumption that the fitted model is correct such that, in particular, the true disturbances are i.i.d. However, this only needs to hold under the null hypothesis, which is appropriate in many situations, e.g. diagnostic tests for incorrect specification.

The block bootstrap implemented, for resampling of stationary dependent data, is the moving blocks method of Kunsch (1989). Integers are drawn with equal probability from 1, …, $n - L$, with replacement, and used to define the beginning of blocks of consecutive observations of length L. Randomly drawn blocks are concatenated, and truncated as required, to form the bootstrap samples of length n.

13.5.6 Stationary Block Bootstrap

This variant of the block bootstrap, due to Politis and Romano (1994), draws blocks with random length as well as random initial observation, where the block lengths are drawn independently from the geometric distribution with parameter $1/L$, such that the mean block

©James Davidson 2015

length is L. For the purpose of drawing the blocks the residual series is "wrapped", such that $\hat{u}_{T+1} = \hat{u}_1$, $\hat{u}_{T+2} = \hat{u}_2$ and so forth.

13.5.7 Wild Bootstrap

The wild bootstrap (see Gonçalves and Kilian 2003, Davidson, Monticini and Peel 2007, Davidson and Flachaire 2008) is robust to neglected heteroscedasticity. If $(\hat{u}_1,...,\hat{u}_T)'$ are the model residuals, the re-sampled variates in this case are $u_t^* = \hat{u}_t \eta_t$ for $t = 1,\ldots,T$ where $(\eta_1,\ldots,\eta_T)'$ is a vector of i.i.d. drawings from a 2-point distribution,

$$\eta_t = \begin{cases} a \text{ with probability } 1/(1+a^2) \\ -1/a \text{ with probability } a^2/(1+a^2). \end{cases} \tag{13.9}$$

The first two moments of the distribution are preserved for any choice of $a > 0$. Setting $a = 1$ is equivalent to a random change of sign, always yields a symmetric distribution while preserving the kurtosis. Setting

$$a = \frac{\sqrt{5}+1}{2} = 1.618 \tag{13.10}$$

sets the skewness equal to that of the original distribution, but doubles the kurtosis. The program prints the Kolmogorov-Smirnov statistic comparing the sample with the bootstrap distribution, as a guide to choosing the best value of a.

13.5.8 Fourier Bootstrap

The Fourier bootstrap applies the fast Fourier transform (FFT) to the residuals. The transformed series is approximately serially independent but heteroscedastic in large sample, with variance at frequency j given by the spectral density at j. A resampling step makes a drawing from this series using the Rademacher wild bootstrap, and then applies the inverse FFT to obtain the bootstrap sample.

This method provides for more general forms of stationary autocorrelation than the sieve-AR bootstrap, but is not suitable for heteroscedastic or non-Gaussian processes. (Note: the CLT implies that the bootstrap series are necessarily Gaussian in large samples, whatever the parent distribution.)

13.5.9 Sieve-AR Bootstrap

Any of the bootstrap methods can be combined with a sieve autoregression to allow for dependent data (see Buhlmann 1997). The residuals are "whitened" by fitting an autoregressive process. The maximum lag order is chosen by default as $[0.6T^{1/3}]$, and can be set manually as an option. The actual lag order, p, is chosen to optimize the Schwarz information criterion in the specified interval. The resampled residuals are then "re-coloured" by the fitted AR(p) in each replication. This procedure can be applied in conjunction with the simple bootstrap or wild bootstrap, and dependence can also be dealt with by a combination of sieve-AR pre-whitening and the blocking or Fourier techniques.

13.5.10 Data Resampling

In this method the data (rather than the residuals) are resampled randomly, with replacement. This is only suitable for cases where the observations are serially independent, hence only for cross-section samples, in general. However, note that all the variables in the model, both dependent and explanatory, are resampled jointly. The range of

©James Davidson 2015

the data set from which the observations are drawn can be different from the estimation sample itself.

13.6 Data Generation

In a simulation run, equations (4.1)+(4.2) or (4.21)+(4.2), are inverted to obtain Y_t from u_t, or $\boldsymbol{Y}_t$ from $\boldsymbol{u}_t$ in the system counterpart (5.1). When equations are specified in the Linear Regression dialog, these are treated for the purposes of simulation as special cases of these dynamic formulations. In conditional heteroscedasticity models, equations (6.1) or (6.2), as appropriate, are used to generate the conditional variances and get u_t from e_t, or (7.1) or (7.2) to get $\boldsymbol{u}_t$ from $\boldsymbol{e}_t$ in the system case. Exogenous variables, denoted as $\boldsymbol{x}_{kt}$ for k = 1,…,6 in the equations, are treated as fixed. To have regressors drawn randomly in repeated runs, as in a Monte Carlo experiment, it is generally necessary to generate them from recursive equations in a multi-equation model, although there is an option in the coded function feature to insert i.i.d.-standard normal data in place of a measured variable – see the Users' Manual Section 1.5 for details.

13.6.1 Dynamic Data Simulation

By default, the dynamics of the specified model are fully reproduced in the simulation procedure. Inversion of the AR, MA and fractional lag components is generally self-explanatory. Inverted lag polynomials are truncated at the start of the available data, or at the truncation point specified under "Dynamic Model Settings" in the Options / ML and Dynamics dialog, if this is set. Unobserved pre-sample values are set to 0, or, in the case of ARCH/GARCH processes, to the appropriate power of the innovation variance (not the unconditional variance, note, since this may be undefined, as in nonstationary cases).

In the cases where Y_t is replaced by a coded function f_{1t} in (4.31), the equation is first solved for f_{1t} from u_t, then the coded function $g_1(...)$ specified in f_{1t} is added to it to generate Y_t. Note that simulations cannot be performed with the "Residual" coding variant in (4.32). However, it is generally possible to create an equivalent model, dedicated to simulation, using the W# reserved name to represent f_{1t} in (4.32). Be careful to note too that $f_{1t} = u_t$ only in the case where none of the programmed model features are specified.

If a pre-sample period is specified – in other words, if some initial observations are excluded from the selected sample – there are two options for setting the initial conditions (see the Options / Simulation and Resampling dialog). In the case "Fixed Presample Data" the actual data set, and estimated residuals, are used. This mode is always adopted for Monte Carlo forecasting and bootstrap calculations. Otherwise it is user-selectable. In the case "Random Presample Data", the simulation run for a stationary process actually starts at date 1 (start of the observed sample), although the simulation is only reported from the specified start date. By choosing a long enough presample period, the presample data can therefore attain their stationary distribution. If a unit root is specified, however, the cumulation starts at the initial date specified, not date 1.

Long memory models pose a special problem because the dependence on presample shocks is potentially large and persistent. If these effects are suppressed, the resulting distributions of partial sums converge to functionals of the "type II" fractional Brownian, having nonstationary increments. Even a very large number of presample lags may fail to reproduce the stationary process whose partial sums converge to the "type I" fractional Brownian motion. An alternative simulation method is provided for this case, which

©James Davidson 2015

generates the stationary process without any requirement to set presample lags – see Davidson and Hashimzade (2009) for the details.

Discrete data models are simulated by making conditional drawings from the distributions specified by the likelihood model, Probit, Tobit or a count data distribution, as the case may be. Note that these drawings may be serially dependent through their dependence on the conditioning variables, but the models are not otherwise dynamic.

Markov switching models are generated by making random drawings from the relevant switch distribution. This source of randomness is additional to the generation of the equation disturbances in each regime, which are drawn in parallel from the specified distributions – switching variances are of course possible.

13.6.2 The Static Bootstrap

If the "Static" option for bootstrap tests is selected, the simulated data are generated by adding the resampled residuals to the fitted values from the estimated model. In the case of static equations in which all explanatory variables are exogenous (i.e. held fixed in the replications) this option has no effect on the simulation procedure, although it should generally run considerably faster than the default simulation procedure. In dynamic equations, the generated data are conditioned on the actual lagged dependent variables, not the generated lagged dependent variables.

In this case the resampled residuals are not corrected for conditional heteroscedasticity, as is done in those models where these effects are modelled, and hence re-introduced as part of the simulation. This means that a resampling method robust to heteroscedasticity must be adopted, either the wild bootstrap or one of the block bootstrap variants.

Note that this simulation option is available only in the context of bootstrap tests. Other applications, such as Monte Carlo exercises, must use the full dynamic simulation method. It is likewise not available for regime-switching models.

13.7 Nonlinear Models

The bootstrap is computationally burdensome when applied to nonlinear models estimated by numerical iteration, which must be repeated in every bootstrap replication. However, note that the starting value for the iterations can always be chosen as a consistent point, the 'pseudo-true' model used to generate the data. Davidson and MacKinnon (1999) show that a relatively small number of iterations can suffice to yield a valid bootstrap distribution that improves on the error in rejection probability (ERP) of asymptotic criteria. The optimization is performed by Newton-Raphson steps, using the inverse Hessian evaluated at the "pseudo-true" point as a fixed metric, so that iterations are extremely rapid. It is difficult to give guidelines for choice of the convergence criterion and maximum iterations that optimally trade off speed and best ERP. Experimentation is recommended.

13.8 Panel Data

In panel data models, the actual observations are simulated using the fitted model, regardless of the transformation used in estimation (individual mean deviations, differencing, etc.)

Residuals are calculated directly from the raw data and put into individual mean-deviation form for resampling. A two-stage resampling method is adopted. For each individual in turn, a random (equal probability) drawing with replacement is made from the set of all individuals. The 'within' variation is then created by resampling the residuals relating to this second individual. Any of the implemented resampling schemes for time series can be

©James Davidson 2015

specified for this second stage. In the case of a fixed effects model, the individual's own mean is then added back. In the case of a random effects model, on the other hand, the mean to be added back is that of the second randomly drawn individual. Note that when individual dummies are included in the model, the individual means should be zeros. The same scheme nonetheless continues to apply in principle.

Alternatively, Gaussian disturbances can be generated. These are assigned the specified 'within' variance, and in fixed-effects models, are augmented by the individual's own residual mean, if different from 0. In random effects models, they are augmented by an independent Gaussian drawing having the specified 'between' variance.

13.9 Subsampling Inference

This is an alternative resampling procedure that involves resampling the actual data, not randomly redrawing residuals. The model is fitted to each of the $T - b_T + 1$ contiguous subsamples of length b_T, which must be chosen by the user as a function of sample size, and must satisfy $b_T \to \infty$ and $b_T / T \to 0$. This is easily accomplished in TSM using the rolling regressions feature. Provided the data are stationary and mixing (asymptotically independent) the distributions so generated, suitably re-normalized, converge to the limiting distributions of the full-sample statistics under very general conditions. Subject to these requirements, the subsampling method is not subject to the specification or estimation errors that can affect the parametric bootstrap. On the other hand, it is not suitable for nonstationary or strongly dependent data. For further information on the properties of subsampling tests, and guidance on the choice of subsample length, see for example Politis Romano and Wolf (1999).

Confidence intervals for parameters are computed using the formulae

$$L = \hat{\theta} - b_T t_1 / T, \; U = \hat{\theta} + b_T t_2 / T \tag{13.11}$$

where t_1 and t_2 are defined as in (13.4), with respect to the 2.5% and 97.5% quantile of the subsampling distributions. Null distributions for the Wald statistics, *t*-values and M statistics are generated in the same way as for the bootstrap. However, LM tests cannot be effectively tabulated by subsampling, even in the case of diagnostic tests, since the restrictions of the null hypothesis are not imposed on the data. Hence, the usual asymptotic *p*-values are always given in the output for these, and likewise for the diagnostic *Q* tests and Jarque-Bera test.

The bias correction feature can be implemented in the same manner as for the bootstrap.

13.10 The Fast Double Bootstrap

The fast double bootstrap is a technique proposed by Davidson and Mackinnon (2002, 2007) that may under certain circumstances reduce the error in rejection probability of a bootstrap test. See also Davidson (2006) for an explanation of the principle. In each bootstrap replication, the model is fitted to the generated data, and these estimated parameters are then used to generate a further bootstrap sample. The model is then fitted to these 'second-generation' data to provide the bootstrap statistics for tabulation.

This technique can minimise the error in rejection probability (ERP) due to errors of estimation, particularly when the statistic in question is not asymptotically pivotal. However, it cannot guard against errors of specification, where the model being investigated is different from the DGP of the observed data. It is most useful as a check on

©James Davidson 2015

the performance of the bootstrap. If the two *p*-values (from the single and double bootstraps respectively) differ substantially, caution is advisable in interpreting the results.

13.11 Warp-speed Monte Carlo for Bootstrap Estimators

Monte Carlo experiments on bootstrap inference can be computationally burdensome, since the total number of estimations and data generation cycles becomes the product of the number of Monte Carlo replications (*K*) and the number of bootstrap replications (*B*). The warp-speed method (see Giacomini et al 2013) collapses these stages together. The method has an affinity with the fast-double bootstrap technique. In each Monte Carlo replication, a *single* bootstrap replication is performed. In other words, the cycle of generating a bootstrap sample using the model with fitted parameters is performed just once, not *B* times. At the end of the experiment, it follows that 2*K* test statistics have been computed: the *K* statistics computed using the Monte Carlo data, and the *K* statistics obtained using the bootstrap samples generated from these first *K* estimations. This latter set are sorted and used to derive an empirical distribution (EDF), which stands in for the bootstrap distribution. A bootstrap *p*-value is generated for each Monte Carlo replication by locating the statistic from the first set in this EDF. In a correctly sized test, the distribution of these *p*-values in simulations of the null hypothesis should be uniform on the unit interval, so that for each $\alpha \in [0,1]$, the proportion not exceeding α should approach α as $K \to \infty$. The ERP is estimated as the difference between α and the actual proportion recorded in the experiment.

As a rule, the number of Monte Carlo replications should be set larger than both *B* and *K* in the comparable conventional experiment, but can be substantially less that *BK*. See Giacomini et al. (2013) for a discussion of cases where this method should work, and cases where it should be used with caution.

©James Davidson 2015

14 Additional Features

14.1 Semiparametric Long Memory

The menu item Setup / Semiparametric Long Memory provides a choice of three nonparametric estimators of the long memory parameter *d*. These methods are not a recommended substitute for maximum likelihood estimation of an ARFIMA(*p*,*d*,*q*) model if there is confidence that the ARMA components are correctly specified, but they impose fewer assumptions about the short-run. The assumption is that the spectrum of the process takes the form

$$f(\lambda) = |1 - e^{-i\lambda}|^{-2d} \; f^*(\lambda) \tag{14.1}$$

where f^* represents the short-range component of the dependence. This is assumed smooth in the neighbourhood of the origin, with $f^*(0)' = 0$. Note the alternative representation

$$f(\lambda) = \lambda^{-2d} g(\lambda) \tag{14.2}$$

where g is likewise assumed smooth at the origin with $g(0)' = 0$.

14.1.1 Geweke Porter-Hudak Log-Periodogram Regression

There two variants of the log-periodogram regression method. The Geweke and Porter-Hudak (1983) estimator (GPH) is implemented with trimming and smoothing options available as proposed in Robinson (1995). The regression

$$Y_k^{(J)} = c + dX_k + U_k^{(J)}, k = L + J, L + 2J, \ldots, M \tag{14.3}$$

is computed, where $X_k = -2\log \sin(\lambda_k/2)$ and

$$Y_k^{(J)} = \log\left(\sum_{j=1}^{J} I(\lambda_{k+j-J})\right), \tag{14.4}$$

$I(\lambda_j)$ is the jth periodogram point, and $\lambda_j = 2\pi j/T$. Here J is fixed, and M and L should diverge with sample size such that $M/T \to 0$ and $L/M \to 0$. No guidelines are available for choices of L and J, and if in doubt, set these to 0 and 1 respectively. GPH suggested a bandwidth $M = O(T^{1/2})$. Hurvich, Deo and Brodsky (1998) showed that the optimal MSE bandwidth is of the form $M = CT^{4/5}$, although the optimal C depends on the unknown short-range components of the spectrum, and may need to be small. They also show asymptotic normality only for $M = o(T^{4/5})$. Some evidence on the relative bias and variance in these cases is given in Davidson and Sibbertsen (2009).

14.1.2 Moulines-Soulier Log-Periodogram Regression

The Moulines and Soulier (1999) 'broad-band' log-periodogram method (MS) sets $M = [T/2]$ but models $\log f^*$ by a Fourier expansion of finite order. Regressors of the form $\cos(j\lambda_k)$ are added to the regression for $j = 1,\ldots,P$, where P should diverge with T but $P/T \to 0$. The optimal P depends on the form of $\log f^*$ and therefore it is difficult to offer guidelines, although see Moulines and Soulier (1999) for discussion of different cases. The obvious procedure with either estimator is to experiment and determine the sensitivity of the estimates to these choices.

14.1.3 Local Whittle ML

The third option is the local Whittle Gaussian maximum likelihood estimator, which maximizes the function

©James Davidson 2015

$$L_T(d) = 2d\frac{1}{m}\sum_{j=1}^{M}\log\lambda_j - \log\left(\frac{1}{m}\sum_{j=1}^{M}\lambda_j^{2d}I(\lambda_j)\right) \tag{14.5}$$

where $\lambda_j = 2\pi j/T$ as before. See Robinson (1995b) and Kunsch (1987). Note that this is the concentrated form of the Whittle log-likelihood function

$$L_T(g,d) = -\frac{1}{M}\sum_{j=1}^{M}\left(\log g\lambda_j^{-2d} + \frac{I(\lambda_j)}{g\lambda_j^{-2d}}\right) \tag{14.6}$$

so that the method treats g in (14.2) as a constant, but by choosing M such that $M/T \to 0$, the method yields consistent estimates under suitable assumptions about the properties of $g(\lambda)$ around 0.

The output reports in each case the point estimate, standard error and the t test of significance of d. Also reported, in the case of the log-periodogram regressions, is a Hausman (1978)-type test for the presence of bias in each estimator. This compares the broadband estimate that would be consistent and asymptotically normal if f^* were constant with an estimator that would exhibit less bias if the null hypothesis was false. In the GPH case, this is a narrow-band estimator with $M = 0.64[T/2]$, where 0.64 is a value chosen to maximize the non-centrality, and in the MS case, it is the MS estimator with the selected P. Both of these statistics are asymptotically normal under H_0.

The resampling inference options can be selected for any of these estimation options. In this case, the resampled series is generated by implementing the sieve-AR bootstrap method for the fractionally differenced data, using the estimated $\hat{d}$ to perform the differencing and subsequent re-integration. This allows a bootstrap implementation of the significance test (t-test) on d, and also the Hausman (1978) bias test. Since the pre-sample shocks are set to zero, the data are created using the correction tem derived by Davidson and Hashimzade (2009). This correction is only implemented for stationary series ($|d| < \frac{1}{2}$) so if $\hat{d} > \frac{1}{2}$, the data are integer-differenced, resampled using $\hat{d} - 1$, then re-cumulated using the observed initial observation to supply the initial condition. Note that the resulting series, after normalization, would converge to Type I fractional Brownian motion.

14.1.4 Specification Tests in Geweke-Porter Hudak Estimation

The bias test of Davidson and Sibbertsen (2009) compares wide and narrow-band GPH estimates. The statistic has an asymptotic standard normal distribution (upper-tail rejection) on the null hypothesis of a pure fractional model and, more broadly, in models without significant bias due to neglected short-run dynamics. The skip-sampling test of Davidson and Rambaccussing (2015) tests the null hypothesis of long memory against the alternative of weakly dependent data. This test is most effective (consistent and asymptotically correctly sized) when a pseudo-p-value is computed as a composite with that of the usual Wald test that the memory parameter d is zero; see the cited article for details.

14.2 Cointegration Analysis

The menu item Setup / Cointegration Analysis gives access to the maximum eigenvalue and trace tests of cointegrating rank of a set of I(1) variables, and also to tests of restrictions on the cointegrating space and MINIMAL analysis.

The context for these tests is the Johansen (1988, 1991)-type cointegrating VAR model

$$\Delta x_t = \Pi(L)\Delta x_{t-1} + \alpha\beta' x_{t-1} + \nu + u_t \quad (m\times 1) \tag{14.7}$$

©James Davidson 2015

where $\Pi(L) = \Pi_1 L + \ldots + \Pi_k L^k$ for $k > 0$, 0 otherwise, and α and β are $m \times s$, and the cointegrating rank of the system is $0 \leq s < m$. One option in this dialog reports selection criteria allowing the best choice of k, the additional lag length for the analysis. See Davidson (2000), among many recent references, for details on the theory and implementation of these tests. The tables of critical values used to compute *p*-value inequalities are taken from Osterwald-Lenum (1992).

Asymptotic chi-squared tests, based on the assumption that the cointegrating rank is known, test restrictions on the Johansen matrix of cointegrating vectors, β of the form

$$\exists\, a \text{ such that } H\beta a = 0 \tag{14.8}$$

where H ($p \times m$) is a known matrix representing linear restrictions, here exclusion restrictions. Thus, the null hypothesis specifies that a vector obeying the restrictions lies in the cointegrating space spanned by β. If so, the subset of variables with non-restricted coefficients βa are cointegrating amongst themselves. See Davidson (1998b) or Davidson (2000) Chapter 16.6, on the theory of these tests.

The MINIMAL test algorithm works through all the possible exclusion tests of this type to identify the subsets of variables that are *irreducibly cointegrating*, that is are cointegrated but which contain no cointegrated subsets. It is shown in Davidson (1998a) that identified structural cointegrating relations must be irreducible, and in any case, the maximum likelihood estimates of irreducible relations are mixed Gaussian, so that valid *t* statistics can be computed. For the irreducible vectors that the MINIMAL procedure picks out, implied point estimates and standard errors are reported, and also for sets of more than two variables, the "omission tests" for each included variable, being the chi-squared statistic obtained by omitting it. Also given are the Phillips-Perron statistics computed from the implied cointegrating residuals.

Assessing this evidence allows a decision to be made whether to treat a particular relation as irreducibly cointegrating. Alternative significance levels can be chosen (10%, 5%, 2.5% or 1%) to give a more or less stringent decision criterion on rejecting cointegration. The "rule of thumb correction" suggested by Monte Carlo experiments modifies the test critical values depending on the degree of over-identification implied by the null hypothesis. See Davidson (1998a) for additional details.

14.3 Automatic Model Selection

The program has two routines for automatic model selection. These work by estimating all members of a specified class of models, and reporting the case which optimizes a chosen model selection criterion. The choices accordingly do not depend on test outcomes, and are path-independent. Note that these routines cannot be used in combination.

14.3.1 ARMA Order Selection

ARMA(p,q) models are estimated for all pairs p and q such that p, q and $p + q$ do not exceed specified bounds. The initial case is $p = q = 0$, and parameters are added sequentially using either current estimates or zeros as starting values, as appropriate. Models must be single equations, but may include any other features with the exception of regime switching.

14.3.2 Regressor Selection

If a model is specified to contain N regressors, there are 2^N alternative models containing subsets of this set. All these models can be estimated in sequence and the case yielding the optimal value of a chosen selection criterion (Akaike, Schwarz or Hannan-Quinn) is

©James Davidson 2015

reported. This feature is available for all models except regime switching, although note that if *N* is large and/or the models nonlinear, the time taken to complete the run could also be large.

14.4 SsfPack State Space Modelling

A state space modelling capability is optionally provided by making use of the 'SsfPack Basic' suite of routines, which is available as a free download for academic and teaching purposes. If SsfPack is installed when the TSM installation is run, access to the package is set up automatically.

The models that can be constructed take the general form

$$\begin{aligned} \alpha_{t+1} &= d_t + T_t\alpha_t + H_t\varepsilon_t \qquad (m\times 1) \\ y_t &= c_t + Z_t\alpha_t + G_t\varepsilon_t \qquad (n\times 1) \end{aligned} \tag{14.9}$$

where y_t is an observed time series, α_t is the unobserved state process and by assumption $\varepsilon_t \sim \text{NID}(0, I)$ $(r\times 1)$. The matrices *T*, *Z*, *G*, and *H* and vectors *d* and *c* are usually constants, but can be made time varying, depending on *t* as shown, by having columns of the data matrix ("regressors") supply the elements at time *t*.

Models that can be cast in this form include ARMA models, unobserved component models featuring local level, local trend, seasonal and cyclical components, and cubic splines. The built in features of SsfPack allow these model components to be assembled automatically and model parameters to be estimated by maximum likelihood, although in this mode the user is limited to univariate models (*n* = 1). It is also possible to formulate the models by direct specification of the matrices, although in this case there is no option to estimate unknown parameters. Making use of the Kalman filter, the equations in (14.9) can be used to estimate the states and disturbances from the sample, to construct forecasts and to generate simulations.

The SsfPack procedures are documented in Koopman, Shepherd and Doornik (1998). This article is distributed with the software in PDF format. Another useful reference is Commandeur, Koopman and Ooms (2011).

14.5 Calculator and Matrix Calculator

TSM features facilities for direct data manipulation and the constructions of statistics `by hand', allowing the implementation of various procedures that are difficult to implement as a menu item.

The scalar calculator provides a text box into which formulae for numerical values can be typed in conventional algebraic notation. If formulae include either one or two algebraic symbols (alphabetical string denoting variables) the resulting functions are graphed, in 2D or 3D plots, respectively. In this case, the user is prompted for upper and lower bounds for each variable. The number of points to be plotted is selectable.

The matrix calculator provides a dialog where moment matrices (mean squares and mean products) can be computed for any pair of sets of variables from the data matrix. The resulting matrices (square and symmetric when the same variable sets are paired) can be named and stored, printed out, and saved in vectorized form in a spreadsheet file. Matrices can also be created and edited directly in a tabular format.

The matrix calculator itself is a text box into which matrix expressions can be entered using conventional notation. Transposes, sums, products (including Hadamard and Kronecker

©James Davidson 2015

products), inverses and generalized inverses, and vectorizations may be constructed and combined in the usual way. The usual scalar functions including determinant, trace and norm can be computed, printed out and also saved as one-dimensional matrices. Eigenvalues and eigenvectors are available, but for symmetric matrices only since complex numbers cannot be handled.

©James Davidson 2015

15 References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, 59, 817-58.

Andrews, D. W. K. (1993) Tests for parameter instability and structural change with unknown change point, *Econometrica* 61, 821-856

Andrews, D. W. K., and J. C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, *Econometrica*, 60, 953-66.

Andrews, D. W. K., and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica*, 62, 1383-1414.

Arellano, M. (2003) Panel Data *Econometrics,* Oxford University Press.

Arellano, M. and S. Bond (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economic Studies* 58, 277-297

Bhargava, A., L. Franzini and W. Narendranathan (1982) Serial Correlation and the fixed effects model. *Review of Economic Studies* 49, 533-549

Bierens, H. J. (1990) A consistent conditional moment test of functional form. *Econometrica*, 58, 1443-1458

Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31, 307-27.

Box, G. E. P. and Cox, D. R. (1964) An analysis of transformations. *Journal of Royal Statistical Society*, Series B, vol. 26, pp. 211-–246.

Box, G. E. P., and D. A. Pierce (1970) The distribution of residual autocorrelations in autoregressive-integrated moving average time series models, *Journal of the American Statistical Association* 65, 1509-26.

Breitung, J., and U. Hassler (2002) Inference on the cointegration rank of fractionally integrated processes, *Journal of Econometrics* 110, 167-185.

Breusch, T. S. and A. R. Pagan (1980) The Lagrange multiplier test and its applications to model specification in econometrics, *Review of Economic Studies* 47, 239-253.

Brown, R.L., J. Durbin, & J.M. Evans (1975) Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society*, Series B 37, 149–163.

Bühlmann, P. (1997) Sieve bootstrap for time series, *Bernoulli* 3, 123-148.

Bunzel, H., N. Kiefer and T. Vogelsang (2001) Simple robust testing of hypotheses in nonlinear models, *Journal of the American Statistical Association* 96, 1088-1096

Cameron, A. C. and P. K. Trivedi (1986) Econometric models based on count data: comparisons and applications of some estimators and tests. *Journal of Applied Econometrics* 1, 29-53.

Commandeur, J. J. F., S. J. Koopman and M. Ooms (2011) Statistical software for state space methods, *Journal of Statistical Software* 41,1-18.

Davidson, J. (1998a) A Wald test of restrictions on the cointegrating space based on Johansen's estimator. *Economics Letters* 59, 183-7.

©James Davidson 2015

Davidson, J. (1998b) Structural relations, cointegration and identification: some simple results and their application. *Journal of Econometrics* 87, 87-113.

Davidson, J. (2000) *Econometric Theory*, Oxford, Blackwell Publishers.

Davidson, J. (2002) A model of fractional cointegration, and tests for cointegration using the bootstrap, *Journal of Econometrics* 110(2) pp187-212

Davidson, J. (2004a) Moment and memory properties of linear conditional heteroscedasticity models, *Journal of Business and Economics Statistics* 22 (1), pp 16-29.

Davidson, J. (2004b) Forecasting Markov-switching dynamic processes, *Statistics and Probability Letters* Vol. 68 (2) pp 137-147.

Davidson, J. (2005) Testing for fractional cointegration: the relationship between government popularity and economic performance in the UK, in *New Trends in Macroeconomics*, eds. C. Diebolt and C. Kyrtsou, Springer Verlag.

Davidson, J. (2006) Alternative bootstrap procedures for testing cointegration in fractionally integrated processes" *Journal of Econometrics* 133 (2) 741-777.

Davidson, J. (2009) When is a time series I(0)?, Chapter 13 of *The Methodology and Practice of Econometrics*, eds. Jennifer Castle and Neil Shepherd, Oxford University Press.

Davidson, J., D. Byers and D. Peel (2006) Support for Governments and Leaders: Fractional Cointegration Analysis of Poll Evidence from the UK, 1960-2004. *Studies in Nonlinear Dynamic and Econometrics* 10.1.

Davidson, J., A. Monticini and D. Peel (2007) Implementing the wild bootstrap using a two-point distribution", *Economics Letters* 96(3), 309-315

Davidson, J. and P. Sibbertsen (2009) Tests of Bias in Log-Periodogram Regression *Economics Letters* 102, 83-86.

Davidson, J. and N. Hashimzade (2009) Type I and type II fractional Brownian motions: a reconsideration, *Computational Statistics and Data Analysis* 53(6) 2089-2106

Davidson, J. and A. Halunga (2013) Consistent tests of functional form in dynamic models. Chapter 2 of *Essays in Nonlinear Time Series Econometrics,* eds. N. Haldrup, P. Saikkonen and M. Meitz (Oxford University Press).

Davidson, J. and D. Rambaccussing (2015) A test of the long-memory hypothesis based on self-similarity *Journal of Time Series Econometrics* 7(2) (2015) 115-142

Davidson, R. and E. Flachaire (2008) The wild bootstrap, tamed at last. *Journal of Econometrics* 146, 162-169.

Davidson, R., and J. G. MacKinnon (1999b) Bootstrap testing in nonlinear models, *International Economic Review* 40, 487–508.

Davidson, R. and J. MacKinnon (2002) Fast double-bootstrap tests of nonnested linear regression models. *Econometric Reviews* 21, 417-427.

Davidson, R. and J. MacKinnon (2007) Improving the reliability of bootstrap tests with the fast-double bootstrap. *Computational Statistics and Data Analysis* 51, 3259-3281.

Davies, R. B. (1977) Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* 64, 247-254.

 ©James Davidson 2015

Dempster, A.P., N.M. Laird, and D.B. Rubin. (1977) Maximum-likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* Series. B., Vol. 39.

Deng, A., and P. Perron (2008) The limit distribution of the cusum of squares test under general mixing conditions. *Econometric Theory*, 24, 809–822

Den Haan, W. J., and A. Levin, (1997) A practitioner's guide to robust covariance matrix estimation. Chapter 12 (pp291-341) of *Handbook of Statistics* 15, North-Holland-Elsevier.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association* 74, 427-431.

Ding, Z., C. W. J. Granger, and R. F. Engle (1993) A Long Memory Property of Stock Market Returns and a New Model, *Journal of Empirical Finance* 1, 83–106.

Doornik, J. A. (1999) *Object-Oriented Matrix Programming Using Ox, 3rd ed.* London: Timberlake Consultants Ltd and Oxford: www.nuff.ox.ac.uk/Users/Doornik.

Elliott, G., T. J. Rothenberg and J. H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64(4), 813-836

Engle, R. F. (2002) Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economic Statistics*, Vol. 20, No. 3, pp.339-350.

Engle, R. F. and K. Kroner (1995) Multivariate simultaneous generalized ARCH. *Econometric Theory* 11, 122–50.

Engle, R. F., and A. D. Smith (1999) Stochastic permanent breaks," *Review of Economics and Statistics*, 81, 553–574.

Feller, W. (1948) On the Kolmogorov-Smirnov limit theorems for empirical distributions, *Annals of Math. Stat.* 19, 177-169.

Fernandez, C., and M. Steel (1998) "On Bayesian modelling of fat tails and skewness," *Journal of the American Statistical Association* 93, 359–371.

Fokianos, K., A. Rahbek and D. Tjøstheim (2009) "Poisson autoregression", *Journal of the American Statistical Association* 04, 1430-1439.

Geweke, J. and S. Porter-Hudak (1983) The estimation and application of long-memory time series models. *Journal of Time Series Analysis* 4, 221–237

Giacomini, R., D. N. Politis and H. White (2013) A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators, *Econometric Theory* 29, 567-589.

Giraitis, L, P. Kokoszka, R. Leipus and G. Teyssière (2003) Rescaled variance and related tests for long memory in volatility and levels, *Journal of Econometrics* 112, 265-294.

Glosten, L., R. Jagannathan, and D. Runkle (1993) "On the relation between expected value and the volatility of the nominal excess return on stocks," *Journal of Finance*, 48, 1779–1801.

Goffe, William L., Gary D. Ferrier, and John Rogers (1994). Global Optimization of Statistical Functions with Simulated Annealing. *Journal of Econometrics*, 60(1/2):65-99.

Gonçalves, S, and L. Kilian. (2003) Bootstrapping autoregressions with conditional heteroscedasticity of unknown form, University of Montreal Working Paper

 ©James Davidson 2015

González, Andrés (2004) "A Smooth Permanent Surge Process", Stockholm School of Economics SSE/EFI Working Paper Series in Economics and Finance No.572

Gonzalo, J., and O. Martinez (2006) "Threshold integrated moving average process (Does size matter? Maybe so)", *Journal of Econometrics* 135, 311-347.

Granger, C. W. J. and T. Teräsvirta (1993) *Modelling Nonlinear Economic Relationships*. Oxford University Press.

Hall, P. (1992) *The Bootstrap and Edgeworth Expansion*, New York, Springer Verlag.

Hamilton, J. D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357-384.

Hamilton, J. D. and R. Susmel (1994) Autoregressive conditional heteroscedasticity and changes in regime, *Journal of Econometrics* 64, 307-333

Hansen, B. E. (1990) Lagrange multiplier tests for parameter instability in nonlinear models. Mimeo at http://www.ssc.wisc.edu/~bhansen/papers/LMTests.pdf

Hansen, B. E. (1992) Testing for parameter instability in linear models. *Journal of Policy Modelling* 14, 517-533.

Hansen, B. E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413-430

Harris, D, B. McCabe and S. Leybourne (2008) Testing for long memory. *Econometric Theory* 24(1), 143-175.

Hauser, M. A. (1999) Maximum likelihood estimators for ARMA and ARFIMA models: a Monte Carlo study *Journal of Statistical Planning and Inference*, 80 (1-2). 229-255

Hausman, J. A. (1978) Specification tests in econometrics, *Econometrica* 48, 1251-1271.

Horowitz, J. L. (2000). The Bootstrap. Chapter for *Handbook of Econometrics* Vol. 5, eds. J. J. Heckman and E. Leamer, North-Holland Elsevier .

Hurvich, C. M., R. Deo and J. Brodsky (1998) The mean squared error of Geweke and Porter-Hudak's estimator of a long memory time series, *Journal of Time Series Analysis* 19, 19-46.

Johansen, S. (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231-54.

Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551-80.

Kiefer, N., T. Vogelsang and H. Bunzel (2000), Simple robust testing of regression hypotheses *Econometrica*, Vol. 68, No. 3. (May, 2000), pp. 695-714.

Kiefer, N. and T. Vogelsang (2002a), Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica*, 70 (5) , 2093-2095.

Kiefer, N. and T. Vogelsang (2002b), Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size *Econometric Theory*, 18, 1350–1366

Kim, C.-J. and C. R. Nelson (1999) *State-space Models with Regime Switching: Classical and Gibbs-sampling Approaches with Applications* MIT Press

Koopman, S. J., N. Shepherd and J. A. Doornik (1998) Statistical lgorithms for models in state space using SsfPack 2.2, *Econometrics Journal* 1, 1-55

 ©James Davidson 2015

Kunsch, H. R. (1987) Statistical aspects of self-similar processes, *Proceedings of 1st World Congress of the Bernoulli Soc.* (Eds. Yu Prohorov and V. V. Sazanov) VNU Science Press, Utrecht 1, 67-74.

Kunsch, H.R. (1989) The jack-knife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217–1241.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin (1992) Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics* 54, 159-178.

Lambert, P. and S. Laurent (2001) "Modelling financial time series using GARCH-type models with a skewed Student distribution for the innovations" Working Paper, .

Li, H. and Maddala, G. S., (1996) Bootstrapping time series models, *Econometric Reviews* 15, 2, 1-115.

Ljung, G. M., and G. E. P. Box (1978). "On a measure of lack of fit in time-series models," *Biometrika*, 65, 297 303.

Lo, Andrew W. (1991) Long-term memory in stock market prices, *Econometrica* 59, 5, 1279-1313

Lütkepohl, H. (2007) *New Introduction to Multiple Time Series Analysis*, Springer

MacKinnon, J. G. (1991) "Critical values for cointegration tests," Ch. 13 in Long-run Economic Relationships: Readings in Cointegration, eds. R. F. Engle and C. W. J. Granger, Oxford, Oxford University Press.

McLeod, A. I. and Li, W. K. (1983) Diagnostic checking ARMA time series models using squared-residual autocorrelations, *Journal of Time Series Analysis* 4, pp. 269--273.

Moulines, E. and P. Soulier (1999) Broad band log-periodogram estimation of time series with long-range dependence. *Annals of Statistics* 27, 1415-1439

Newey, W. K., and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-8.

Newey, W. K., and K. D. West (1994). Automatic lag selection in covariance matrix estimation, *Review of Economic Studies* 61, 631-653

Nelson, D. B. (1991) Conditional heteroscedasticity in asset returns: a new approach. *Econometrica* 59, 347–70.

Nyblom, J. (1989) Testing for the constancy of parameters over time. *Journal of the American Statistical Association* 84, 223-230

Osterwald-Lenum, M. (1992) A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics. *Oxford Bulletin of Economics and Statistics* 54, 461-72.

Phillips, P. C. B. and P. Perron (1988) Testing for a unit root in time series regression, *Biometrika* 75, 335-346

Politis, D. N. and J. P. Romano (1994) The Stationary Bootstrap. *Journal of the American Statistical Association* 89, 1303-1313.

Politis, D. N., J. P. Romano and M. Wolf (1999) *Subsampling*, Springer Verlag.

Psarakis S., and J. Panetaros (1990) The folded *t* distribution, *Communications in Statistics, Theory and Methods* 19 (7) 2717-2734

 ©James Davidson 2015

Priestley, M. B., (1988) *Non-Linear and Non-Stationary Time Series Analysis*. London: Academic Press.

Robinson, P. M. (1994) Semiparametric analysis of long memory time series, *Annals of Statistics* 22, 1, 515--539.

Robinson, P. M. (1995a) Log-periodogram regression of time series with long-range dependence, *Annals of Statistics* 23, 1048–1072.

Robinson, P. M. (1995b) Gaussian semiparametric estimation of long-range dependence, *Annals of Statistics* 23, 1630–1661.

Robinson, P. M. and I. N. Lobato (1998) A nonparametric test for I(0), *Review of Economic Studies* 65(3), 475-495.

Said, E. S. and D. A. Dickey (1984). Testing for unit roots in autoregressive-moving average models of unknown order, *Biometrika* 71, 599-607.

Saikkonen, P. (1991). Asymptotically efficient estimation of cointegration regressions, *Econometric Theory*, 7, 1–21.

Stock, J. H and Watson, M. W. (1993) A simple estimator of cointegrating vectors in higher order integrated systems, *Econometrica,* 61, 783-820.

Subba Rao, T. (1981) On the theory of bilinear models, *Journal of the Royal Statistical Society* B 43, 244-245.

Tanaka, k., (1999) The nonstationary fractional unit root. *Econometric Theory* 15, 549-582.

Teräsvirta, T.: 1998, Modeling economic relationships with smooth transition regressions, in A. Ullah and D. E. Giles (eds), *Handbook of Applied Economic Statistics*, Dekker, New York, pp. 507–552.

Tong, H., (1990) *Non-Linear Time Series. A Dynamical System Approach.* Oxford: Clarendon Press.

Vuong, Q. H. (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57, 307-333.

White, Halbert (1980) A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity, *Econometrica* 48, 817–838.

 ©James Davidson 2015

Index

2SLS, 7
3SLS, 7, 8
ADF test. *See* augmented Dickey-Fuller test
Akaike Information Criterion, 36
Andrews parameter stability test, 52
APARCH, 26, 27, 28
ARCH, 13
ARFIMA, 13, 15, 57, 67, 75
ARIMA, 13, 14
ARMA, 13, 14, 15, 16, 26, 30, 37, 41, 57, 67, 69, 75, 76
asymmetric error correction, 18, 27
augmented Dickey-Fuller test, 7, 8, 56
autoregressive moving average, 13
bandwidth, 8, 41, 44, 46, 48, 56, 57, 67, 75
Bartlett kernel, 46, 48, 75
BEKK, 29, 30
Bierens, 72
Bierens test, 52
bilinear model, 15, 16
binary dependent variable, 21, 28
block bootstrap, 61
Bollerslev, 26, 72
bootstrap test, 56
Box-Cox transformation, 19
Box-Pierce, 37
Breitung, 56
Breusch-Pagan test, 12
calculator, 70
chi-squared probit, 21
Chow test, 7, 38
coded formulae, 18
cointegration, 8
common factor test, 50
conditional moment test, 7, 49
consistent specification tests, 52
count data, 22
cubic polynomial, 18
cusum of squares, 54
data generation, 63
data resampling, 62
data transformations
  panel data, 9
DCC, 29, 30
delta method, 15
discrete autoregression, 22
dummy variables
  panel data, 9
Durbin-Watson test, 7, 12, 37
Durbin-Wu-Hausman, 37
dynamic specification tests, 53
ECM, 13, 16, 18, 25
EGARCH, 26, 27
Elliott-Rothenberg-Stock tests, 56
EM algorithm, 45
empirical distribution function, 58
equilibrium relations, 17, 20, 24, 25
error correction model, 17, 20
error dynamics, 15
ESTAR, 18
explained switching, 32
exponential smooth transition, 18
ex-post forecasts, 37
fast double bootstrap, 65, 66
FIEGARCH, 28
FIGARCH, 27
filter probabilities, 31, 35
FIML, 13, 23, 44, 45
fixed effects, 10
Fourier bootstrap, 62
fractional cointegration, 18, 24, 25, 73
fractional difference, 14
full information maximum likelihood. *See* FIML
FVECM, 24, 25
GARCH, 13, 26, 27, 28, 30, 33, 38, 42, 63, 76
GARCH-M, 26
Gaussian distribution, 61
generalized cointegration, 25
generalized error distribution (GED), 58
Generalized Error Distribution (GED), 42, 45
generalized least squares, 10
generalized method of moments, 11, 44
Geweke Porter-Hudak, 67
GMM. *See* generalized method of moments
HAC, 41, 44, 46, 47, 48, 49, 50
Hamilton, 32, 33, 39, 75
Hannan-Quinn criterion, 36
Harris-McCabe-Leybourne test, 56
Hausman test, 12
Hessian, 46, 64

©James Davidson 2015

heteroscedasticity, 26, 28, 44, 45, 50
HYEGARCH, 28
HYGARCH, 27
IGARCH, 27
impulse responses, 39
inequality constraints, 15
information matrix, 51
Information Matrix, 46
instrumental variables, 11, 41
integration order, 55
intercept types, 14
IV. *See* instrumental variables
Jacobian matrix, 49
Jarque-Bera test, 11, 36, 60, 65
Johansen, 68, 69, 72, 75
Kalman filter, 70
Kim's algorithm, 35
KPSS, 57
KPSS test, 55
KVB, 46, 48
lag distribution, 14
lag polynomials, 13, 14, 18, 23, 26
Lagrange multiplier test, 7, 49
least generalized variance, 43
Least Generalized Variance, 44
least squares, 10, 41
LGV. See least generalized variance
Ljung-Box, 37
loadings coefficients, 17, 24
logistic transformation, 15
logit, 21, 28, 35, 58
log-periodogram regression, 67
MA coefficients, 14
Markov switching, 13, 31, 32, 33
matrix calculator, 70
maximum likelihood, 10, 36, 41, 42, 51, 61, 67, 69, 76
McLeod-Li, 37
MINIMAL, 68, 69
moment test, 49
Monte Carlo experiments, 40, 58, 60, 61, 63, 69, 72, 75
Monte Carlo forecasts, 40
Moulines-Soulier, 67
negative binomial, 43
NegBin I, 22
NegBin II, 22
neglected ARCH, 50
Newey-West, 44, 46
nonlinear autoregression, 17
nonlinear error correction, 18
nonlinear functional form, 50
Nyblom-Hansen, 51
ordered logit, 21, 28, 43
ordered probit, 21, 28
ordered Probit, 43
ordinary least squares, 7, 10
Osterwald-Lenum, 69, 76
Ox Coding, 20
panel data, 9
parallel processing, 58
parametric bootstrap, 59, 65
Parzen kernel, 46
percentile-*t*, 60
Phillips-Hansen, 8
Phillips-Perron test, 7, 8
plug-in bandwidth, 46, 48
Poisson, 22, 28, 43, 58
polynomial distributed lags, 15
pre-whitening, 47, 48
probit, 21, 28, 35, 42, 58
Q test, 37, 60, 65
Quadratic-Spectral kernel, 46
R/S test, 55
random effects, 10
R-bar-squared, 36
regressor types, 7
residual Aautocorrelation, 50
Robinson-Lobato test, 55
Saikkonen, 8
Sargan, 37
Schwarz information criterion, 36, 62
score-contribution tests, 53
seemingly-unrelated regressions, 7
SETAR, 33
sieve-AR bootstrap, 62
simulation, 63
simultaneous equations, 23
skewed Student's *t*, 41
smooth permanent surge, 17
smooth transition, 33
smoothed probabilities, 31
SsfPack, 70
stable distribution, 61
state space, 70
step responses, 39
STIMA, 16
Stock and Watson, 8
STOPBREAK, 16, 17
structural dynamics, 15
Student's *t*, 41, 44, 57
Subba Rao, 16, 77

©James Davidson 2015

subsampling, 65, 76
Sup-F, 55
sup-Wald, 34
SUR, 7, 12
TAR, 32, 33
transition probabilities, 31, 33
trend dummy, 14, 38
Tukey-Hanning, 46
two-stage least squares, 7
type I fractional Brownian motion, 63, 68
Types. *See* regressor types
V/S test, 55
VAR, 7, 23, 32, 47, 48, 68
VECM, 24
Vuong test, 54
Wald tests, 7
warp speed Monte Carlo, 66
White test, 50
White-Eicker, 44
Whittle, 42, 45, 46, 67, 68
wild bootstrap, 62
zero-inflated, 22, 43

©James Davidson 2015